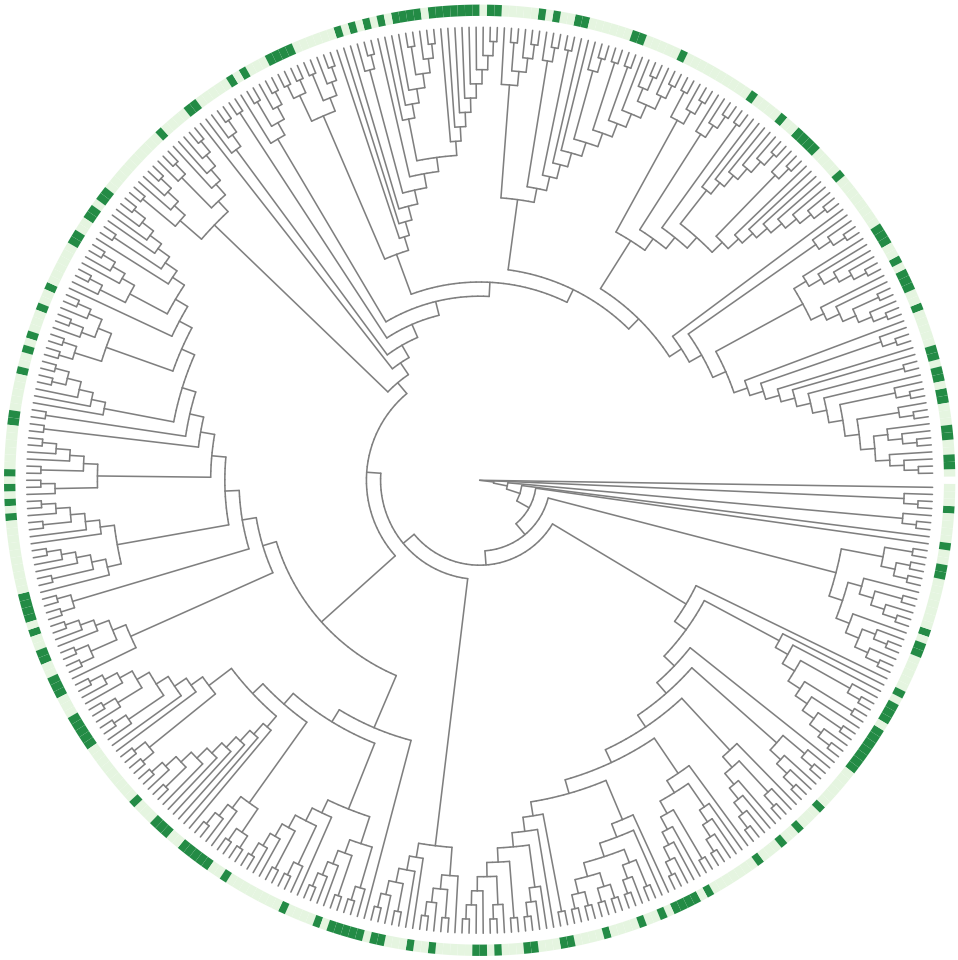
Evolutionary patterns of salt tolerance in angiosperms



Camile Sofia Moray

March 2015

A thesis submitted for the degree of Doctor of Philosophy of The Australian National University.

Declaration

I declare that this thesis is my original work. I am the main contributor to the data collection, analysis, and writing of all the presented work. All chapters have been co-authored with my primary supervisor Lindell Bromham. Bibliographic details, including a list of co-authors are presented on the title page of each chapter. The text has not been altered from published chapters or from chapters that have been submitted for publication. Any structural differences between chapters reflect requirements from the relevant journal/publisher. None of the work in this thesis has been submitted towards another degree.

Camile Sofia Moray

June 2016

Acknowledgments

I am incredibly grateful to have had the opportunity to complete my PhD in such a beautiful country, in a fantastic university, and with an amazing lab group. The last four years have perhaps been the most influential in my life, both professionally and personally. I have gained so many new skills and fostered so many great relationships that I will carry with me throughout my life... along with a few wrinkles and RSI.

Firstly, I would like to thank my primary supervisor, Lindell Bromham, who has supported me through many of life's ups and downs encountered during my PhD. Thank you for giving me this opportunity and for everything you have taught me. Thank you for creating an open-minded, relaxed, and sometimes chaotic lab environment! Thank you for letting me be your right hand **wo**man in organizing and managing the Network for Women in Biology. I would also like to thank my co-supervisors Rob Lanfear and Marcel Cardillo, who are awesome and have always been there to help out and give advice.

Studying in the EEG division has been an excellent experience. The department is filled with so many amazing people, and it's been a pleasure getting to know and spend time with you. In particular, I would like to thank my lab mates for sharing so many laughs and always being there to chat over coffee and cake.

Thank you to my friends and family, especially my parents, who support me even though they hate that I'm on the other side of the world. Finally, I would like to thank my partner, Trevor. Meeting you has been by far the most significant event of my PhD. I'm so grateful for the years we've shared so far, and I look forward to many more.

The work presented in this thesis was supported by an APA(I) scholarship from the Australian Research Council. I would also like to thank the reviewers of this thesis for their time and helpful comments.

Abstract

Global food production is threatened by increasing land salinization triggered by climate change, land clearing, and irrigation. Salinity is toxic to most plants, including most crop species. A tremendous research effort has focused on understanding how a rare set of naturally salt tolerant plants, halophytes, are able to cope with soil salinity, as a model for producing salt tolerant crops. One largely unexplored area of research is the evolution of salt tolerance. Previous studies show that salt tolerance has evolved multiple times across the angiosperms, but little is known about the patterns and processes that underlie the evolution of salt tolerance. In my thesis I addressed several questions relating to the evolution of salt tolerance in angiosperms using a broad-scale, macroevolutionary approach.

I first used taxonomic and phylogenetic comparative techniques to assess the evolutionary patterns of salt tolerance in angiosperms. I found that over one-third of angiosperm families contain halophytes and that salt tolerance appears to have evolved hundreds of times in the angiosperms. In over half of the family phylogenies analyzed, salt tolerance appeared evolutionarily labile: the origins of salt tolerance were scattered across phylogenies and generally gave rise to only one or a few halophytes.

I also explored the association between salt tolerance and another trait associated with anthropogenic environmental change, heavy metal hyperaccumulation: the ability to accumulate high concentrations of heavy metals/metalloids. Taxonomic and physiological similarities suggest that salt tolerance may be associated with hyperaccumulation. I test the suggested relationship between these abilities using taxonomic and phylogenetic analyses. Significantly more angiosperm families contain both halophytes and hyperaccumulators and significantly more species are reported as both halophytes and hyperaccumulators than expected given the rarity of each trait. In several families, halophytes and hyperaccumulators are more closely related than expected if the two traits evolved independently. These results support the observation that salt tolerance and heavy metal hyperaccumulation are associated in angiosperms.

Prolonged or repeated exposure to salinity can cause oxidative stress that may lead to increased mutation rates. These mutations may lead to increased substitution rates in halophytes compared to non-salt tolerant relatives. We tested this idea by comparing

DNA sequences of multiple genes from the chloroplast, mitochondrial, and nuclear genomes from several halophytes with their non-salt tolerant relatives. We found that halophytes have significantly increased total substitution rates compared to their non-salt tolerant relatives in mitochondrial genes. This finding provides evidence that environmental factors may be associated with molecular rates.

The goal of developing salt tolerant crops has proved incredibly difficult, which may be partly due to loss in genetic variation associated with domestication. Yet several studies suggest that domesticated animals and plants may have increased rates of molecular evolution, which could lead to increased variation. We test whether domesticates have consistently increased rates of molecular evolution by comparing the mitochondrial genomes of domesticated mammals and birds to their wild relatives. While a few domesticates exhibited higher rates, in general we found no consistent difference in mitochondrial rates of domesticated animals compared to their wild relatives.

Table of contents

Chapter 1

Introduction

Sections of this chapter are reproduced from my contribution to a published book chapter: “Preliminary analysis”, Figure 1 and Table 1

Published: CH Saslis-Lagoudakis, **C Moray**, and L Bromham. 2014. Evolution of salt tolerance in angiosperms: a phylogenetic approach. In: N. Rajakaruna, R. S. Boyd, & T. B. Harris (Eds.), Plant ecology and evolution in harsh environments. Nova Science Publishers, Hauppauge, NY, pp. 77–95.

Background

Land salinity: a growing threat

Population expansion, agricultural development, and industrialization have caused rapid anthropogenic environmental change across the globe. One particularly widespread problem associated with anthropogenic environmental change is increasing land salinity. Common practices like land clearing and irrigation can increase the amount of water-soluble salts near the surface of the soil, which can be accessed by the root systems of plants. Soil salt is toxic to most plants because it impedes the osmotic potential for water uptake and triggers stress responses within the plant, leading to impaired growth and development, and often death (Flowers and Yeo 1986). It is estimated that between 7-10% of global land is currently salt-affected (Flowers and Yeo 1995; Ruan et al. 2010). In Australia soil salinity is even more widespread, affecting up to 30% of land area (Rengasamy 2006). The effects of land salinization are only predicted to increase, while the difficulty in accessing fresh water also increases, and the rise of sea levels threatens to increase the salinity of coastal areas (Schofield and Kirkby 2003; Rozema and Flowers 2008; Ruan et al. 2010).

Land salinity poses a particular threat to agricultural productivity and sustainability, as the vast majority of crop species cannot grow on salt-affected soils (Rozema and Flowers 2008). One-third of the global food supply is produced on irrigated land, so although only 15% of agricultural land is irrigated, a large portion of productivity is at risk from increasing land salinization (Munns 2002; 2005). By 2050 it is estimated that 50% of arable land will be salt affected (Wang et al. 2003). Yet while land salinity will progressively impede agricultural productivity, the demand for food is also quickly growing: it is predicted that by 2050 food production will need to increase by 70% to feed a global population estimated to reach over 9 billion people (FAO, 2011).

One way to improve agricultural productivity in the wake of increasing land salinization is to develop salt tolerant crops that can grow on salt-affected agricultural land as well as on naturally saline areas, which could be used to sustain and expand crop production (Shannon 1985; Flowers and Yeo 1995; Munns et al. 2006; Rozema and Flowers 2008; Witcombe et al. 2008; Tester and Langridge 2010; Ashraf and Foolad 2012; Panta et al.

2014; Ventura et al. 2015). To breed or engineer salt tolerant crops, researchers have turned to halophytes, naturally occurring salt tolerant plants, to understand how plants can cope with soil salinity. Halophytes are rare amongst plants, representing between 1-2% of plant species, most of which are angiosperms (Glenn et al. 1999; Flowers and Colmer 2008). Halophytes like quinoa (*Chenopodium quinoa*: Gómez-Pando et al. 2010), pearl millet (*Pennisetum glaucum*: BOSTID 1990; Jaradat 2003), and samphire (several species in Salicornioideae, Amaranthaceae: Ventura and Sagi 2013) have been proposed as alternative crops that can produce adequate yields on saline soils and/or can be irrigated with saline or sea water (Panta et al. 2014). But a much larger researcher effort is dedicated to using halophytes as a model to understand how plants can tolerate salinity, and how we can use this knowledge to breed and engineer salt tolerant plant crops. Although researchers have made significant advances in increasing salt tolerance of some major crops (Flowers and Yeo 1995; Ashraf and Foolad 2012; Cheeseman 2014), breeding and engineering salt tolerant crop varieties is an on-going research challenge.

Mechanisms of salt tolerance in plants

One explanation for the difficulty in breeding salt tolerant crops is that salt tolerance is a very complex and diverse trait. Soil salinity makes it more difficult for plants to access water, in some ways mimicking drought conditions (Munns 2002). Salt is also toxic to most plants, and halophytes must control the amount of salt taken up and accumulated in tissues and/or mitigate the effects of salinity within the tissues (Cheeseman 1988; Flowers and Colmer 2008). Salt tolerance in plants involves a number of physiological and sometimes anatomical modifications, and not all halophytes use the same strategies to tolerate salinity. Common physiological mechanisms include 1) ion selectivity (e.g., selecting K^+ over Na^+) and exclusion, which minimizes the amount of salt taken up by the plant; 2) vacuolar compartmentalization, which mitigates the effect of salinity within tissues by storing harmful ions in the vacuole; 3) osmoregulation, correcting osmotic potential to maintain adequate water uptake; and 4) the production of compatible solutes, osmoprotectants, and activation of antioxidant systems to mitigate effects of salinity in cells and protect structures from oxidative damage induced by salinity (Flowers et al. 1977; Greenway and Munns 1980; Munns and Tester 2008; Flowers and Colmer 2008). Some halophytes isolate excess salinity into aerial tissues

and then shed them to remove salt from the plant (Albert 1975), and others use dormancy to avoid germination in particularly harsh conditions (Cao et al. 2014). Some halophytes also have salt glands, hairs, and bladders that can excrete salt (Waisel 1972; Flowers et al. 1977; Ball 1988; Shannon 1997).

Defining salt tolerance in plants

In addition to differences in the mechanisms of salinity tolerance between halophytes, salt tolerance is also a continuous trait that can vary among species: some halophytes can tolerate brackish waters while others can live in salt lakes with salt concentrations that are higher than seawater (Flowers and Colmer 2008). Because salt tolerance exists on a spectrum and can result from different modifications, there are several definitions for identifying and classifying halophytes in the literature. Several studies have used the standard that halophytes are plants that can live and reproduce in soils with ca. 80mM NaCl concentrations or higher (Aronson 1989; Menzel and Lieth 2003), though others have assembled a more restricted lists of species under a ca. 200mM threshold (often called “euhalophytes”: Flowers and Colmer 2008; Flowers et al. 2010). However, these standards can only be accurately verified in controlled experiments, meaning that a limited number of species have been identified as halophytes based on experimental data.

Broader knowledge on the diversity and distribution of halophytes comes from regional field surveys (e.g., Guvensen et al. 2006; Öztürk et al. 2008; Zhao et al. 2011), which are reliant on local soil measurements and observations of local habitat, often splitting halophytes into environmental categories (e.g., xerohalophytes and hydrohallophytes). Although these surveys provide information on a diverse set of habitats, many of the species found in saline areas may not be halophytes in terms of the experimental standards outlined above: for example, some individuals may grow on high sand dunes with less exposure to salinity, and some may only be tolerant to occasional exposure to salinity, such as salt spray. Some species also experience salt acclimation, a type of low salt tolerance gained within an individual plant (e.g., Djanaguiraman et al. 2006; Pandolfi et al. 2012). In this thesis, I aim to investigate the evolution of salt tolerance across the angiosperms. I compile a list of halophytes from the literature to use in broad-scale analyses, including species identified as halophytes in field surveys. Although this approach may include some species that are not halophytes by laboratory

standards, using the rich evidence available in regional surveys is the best way to capture the maximum amount of global diversity of halophytes. Specific information on the list of halophytes used in this thesis can be found in Chapters 2 and 3.

The evolution of salt tolerance

When we look at the phylogenetic relationship between taxonomic orders of angiosperms that have halophytes, it appears that salt tolerance has evolved a number of times independently in the history of the angiosperms (Flowers et al. 1977; 2010). Although they are rare among plants, halophytes are found in a large set of taxonomic families and represent a variety of different life forms (i.e., herbs, shrubs, trees, etc.). Halophytes occupy a range of habitats worldwide (Flowers et al. 1977; Menzel and Lieth 2003), including arid regions and deserts (Özturk et al. 2008), salt marshes and lakes (Pennings and Callaway 1992), and mangrove forests (Ball 1988). Also, because halophytes have adapted to diverse habitats, they often have tolerances to other abiotic stresses like waterlogging (Ball 1998; Colmer and Flowers 2008; Bennett et al. 2009) and drought (Munns 2002; 2011).

The relationship between halophytes at finer taxonomic levels has been explored in relatively few groups. In the sea grasses (Les et al. 1997) and in Amaranthaceae (Kadereit et al. 2012), groups that contain a large proportion of halophytic taxa, salt tolerance appears to have evolved only a few times, early in the history of these lineages, defining large clades of halophytes. One recent study on the grass family, Poaceae, found a strikingly different pattern: that among nearly 3000 grass species, salt tolerance has evolved over 70 times independently, usually near the shallow nodes (tips) of the phylogeny, and that origins of salt tolerance generally give rise to only one or a few halophytes (Bennett et al. 2013). The patterns observed in the sea grasses and Amaranthaceae and in the grasses suggest very different modes of salt tolerance evolution in angiosperms and prompt a broad range of questions including 1) how many times has salt tolerance evolved in angiosperms? 2) if salt tolerance involves many complex physiological/or anatomical modifications, how and why has it evolved so many times in angiosperms?, and 3) if salt tolerance has evolved many times in Poaceae, the family that contains many major crop species, why has it been so difficult to produce salt tolerant crops?

The incredible amount of research on identifying and understanding salt tolerance in a wide range of species provides an exciting backdrop for broad scale, macroevolutionary studies (Bromham 2015). In my thesis I address some of the questions related to the evolution of salt tolerance in the angiosperms in a macroevolutionary context, using phylogenetic comparative methods. Each of the many origins of the evolution of salt tolerance in angiosperms can be treated as an independent observation of the circumstances associated with the evolution of this complex trait. By placing data on salt tolerant species within a phylogenetic context, we can paint a picture of the general features associated with how salt tolerance evolves across a wide range of genetic and ecological backgrounds, and identify common traits or conditions that are associated with these events.

As a preliminary analysis for my thesis, I contributed to a book chapter on the evolution of salt tolerance (Saslis-Lagoudakis et al. 2014). I conducted an analysis on the taxonomic and phylogenetic distribution of known halophytes among angiosperms to address two questions. Firstly, using a published list of halophytes, we estimated the proportion of halophytes among angiosperm families to explore the taxonomic distribution of salt tolerance across angiosperms. Secondly, we examined the phylogenetic distribution of families with halophytes to determine how many times salt tolerance has evolved among angiosperm families and whether closely-related families shared similar proportions of halophytes. The following section ("Preliminary analysis", Table 1, Figure 1) is taken directly from my contribution to Saslis-Lagoudakis et al. (2014).

Preliminary analysis

Taxonomic distribution of halophytes

As a preliminary analysis, we explored the distribution of known halophytes across angiosperm families to investigate if halophytes are distributed randomly across angiosperms. We first recorded the angiosperm families recognised by the APG III and the number of species estimated in each family (http://www.mobot.org/MOBOT/research/APweb/). We then found the number of known halophyte species in each angiosperm family recognised by the APG III, based on a published list of halophytes

(Menzel and Lieth 2003). This list provides approximately 2,600 names of plant species reported as halophytes in published studies based on ecological, physiological and anecdotal data (Menzel and Lieth 2003). Although no published list of halophytes will be complete, due to poor knowledge of salt tolerance in certain families and geographical regions, we believe this is the most extensive published database of known halophytes.

We found the accepted name of each halophytic species in that list by searching The Plant List (2010) with the package "taxonstand" (Cayuela et al. 2012) in the program R (R Core Team 2014). We then allocated each accepted halophyte species to its respective family using the taxonomic name resolution service (TNRS; Boyle et al. 2013). Using this method of estimation, we identified 1,653 halophytic species (Table 1). Based on this survey, we found that halophytes are distributed in 117 families and 34 orders. As expected based on previous studies (Flowers et al. 1977; 2010), many of the families with the highest proportions of halophytes (Table 1) come from the orders Alismatales (including sea grasses) and Caryophyllales (including chenopods). However, there are several families with relatively high proportions of halophytes within the orders Malphigiales, Fagales, and Zygophyllales.

Phylogenetic distribution of halophytes

The distribution of halophytes among taxonomic groups shows that halophytes are found in at least a quarter of angiosperm families. However, we cannot assume that the 117 families with halophytes evolved salt tolerance independently. To estimate the number of origins of salt tolerance across angiosperm families, we carried out a phylogenetic investigation.

For this investigation we used the largest published tree of angiosperms, which contains over 56,000 angiosperm taxa and was constructed from publicly available sequences for six chloroplast and nuclear DNA markers (Smith et al. 2011). From this phylogenetic tree, we extracted a family-level phylogenetic tree, selecting one representative species for each family, randomly choosing between those species with the most sequence data in the alignment. We did not estimate branch lengths for this analysis, and used a phylogenetic tree with all branch lengths set to 1. We used the same list of halophytes described in the taxonomic analysis above, finding the accepted names of the species in

a published list (Menzel and Lieth 2003) according to The Plant List (2010), and using the TNRS (Boyle et al. 2013) to find family affinities. Using a parsimony ancestral state reconstruction method in Mesquite (Maddison and Maddison 2006), we estimated that salt tolerance has evolved independently at least 59 times in the family-level phylogeny of angiosperms (Figure 1).

Although these origins are more prominent in some clades than others, they are dispersed on the phylogeny, with many close to the tips of the family-level tree, so are shared by only one or few families (Figure 1). Further, we explored the phylogenetic distribution of halophyte proportion within each family (Table 1). In Figure 1, we coloured the tips of the phylogeny according to halophyte proportion. We found that families with the highest proportion of halophytes do not appear to be clustered on the angiosperm family tree, but they are sometimes related to families with lower proportions of halophytes (Figure 1).

Of course, based on this result only, we cannot claim there have only been 59 origins of salt tolerance during the evolutionary history of angiosperms. Our analysis is at the family level and, although some families rarely lose salt tolerance (e.g., chenopods, Kadereit et al. 2012 and sea grasses, Les et al. 1997), salt tolerance can be gained several times within a single family. For instance, in the Poaceae, which represent a single tip in our phylogenetic tree (Figure 1), we have identified over 70 origins of salt tolerance (Bennett et al. 2013). Therefore, if we expand our analysis to more shallow taxonomic levels, we expect that the number of estimated origins will only increase. However, it is not clear whether the labile evolutionary pattern of salt tolerance in the grasses is common across many families, or whether the factors driving salt tolerance evolution vary widely across lineages.

Table 1: Estimates for number and percentage of halophytes for 117 families recognised by APG III containing at least one known halophyte. Family names, orders, and estimated species numbers were taken from the APG website version 13 (http://www.mobot.org/MOBOT/ research/APweb/). Number of halophytes was derived from the set of accepted halophyte species included in Haloph v2 (Menzel and Lieth 2003) based on The Plant List (The Plant List 2010), and their respective family affinities according to the Taxonomic Name Resolution Service (Boyle et al. 2013). We highlight families with more than 50 species in bold. Families are ranked alphabetically by the order to which they belong.

Order	Family	Number of Species	Number of Halophytes	Percentage of Halophytes
Alismatales	**Alismataceae**	88	1	**1.14**
Alismatales	Cymodoceaceae	16	12	75.00
Alismatales	**Hydrocharitaceae**	116	13	**11.21**
Alismatales	Juncaginaceae	15	1	6.67
Alismatales	Posidoniaceae	9	3	33.33
Alismatales	**Potamogetonaceae**	102	7	**6.86**
Alismatales	Ruppiaceae	6	1	16.67
Alismatales	Zosteraceae	14	14	100.00
Apiales	**Apiaceae**	3780	9	**0.24**
Apiales	**Araliaceae**	1450	3	**0.21**
Arecales	**Arecaceae**	2361	29	**1.23**
Asparagales	**Amaryllidaceae**	1605	4	**0.25**
Asparagales	**Asparagaceae**	2480	9	**0.36**
Asparagales	**Iridaceae**	2025	4	**0.20**
Asparagales	**Xanthorrhoeaceae**	900	1	**0.11**
Asterales	**Asteraceae**	23600	117	**0.50**
Asterales	**Calyceraceae**	60	2	**3.33**
Asterales	**Campanulaceae**	2380	1	**0.04**
Asterales	**Goodeniaceae**	430	4	**0.93**
Brassicales	Bataceae	2	2	100.00
Brassicales	**Brassicaceae**	3710	21	**0.57**
Brassicales	**Capparaceae**	480	1	**0.21**
Brassicales	**Cleomaceae**	300	3	**1.00**
Brassicales	**Resedaceae**	75	1	**1.33**
Brassicales	Salvadoraceae	11	1	9.09
Caryophyllales	**Aizoaceae**	2035	36	**1.77**
Caryophyllales	**Amaranthaceae**	2275	393	**17.27**
Caryophyllales	Anacampserotaceae	32	1	3.13
Caryophyllales	Basellaceae	19	2	10.53
Caryophyllales	**Cactaceae**	1866	8	**0.43**
Caryophyllales	**Caryophyllaceae**	2200	17	**0.77**

Order	Family	Number of Species	Number of Halophytes	Percentage of Halophytes
Caryophyllales	Didiereaceae	16	2	12.50
Caryophyllales	**Frankeniaceae**	90	15	**16.67**
Caryophyllales	Halophytaceae	1	1	100.00
Caryophyllales	**Molluginaceae**	87	1	**1.15**
Caryophyllales	**Montiaceae**	226	3	**1.33**
Caryophyllales	**Nyctaginaceae**	395	7	**1.77**
Caryophyllales	**Plumbaginaceae**	836	28	**3.35**
Caryophyllales	**Polygonaceae**	1110	22	**1.98**
Caryophyllales	**Portulacaceae**	70	5	**7.14**
Caryophyllales	Sarcobataceae	2	1	50.00
Caryophyllales	Stegnospermataceae	3	1	33.33
Caryophyllales	Talinaceae	27	2	7.41
Caryophyllales	**Tamaricaceae**	90	28	**31.11**
Celastrales	**Celastraceae**	1400	8	**0.57**
Commelinales	Pontederiaceae	33	1	3.03
Cucurbitales	**Cucurbitaceae**	960	3	**0.31**
Dilleniales	**Dilleniaceae**	355	1	**0.28**
Ericales	**Ebenaceae**	548	2	**0.36**
Ericales	**Ericaceae**	3995	1	**0.03**
Ericales	**Lecythidaceae**	310	4	**1.29**
Ericales	**Primulaceae**	2590	10	**0.39**
Ericales	**Sapotaceae**	1100	2	**0.18**
Ericales	Tetrameristaceae	5	1	20.00
Fabales	**Fabaceae**	19500	113	**0.58**
Fabales	Surianaceae	8	1	12.50
Fagales	**Casuarinaceae**	95	9	**9.47**
Gentianales	**Apocynaceae**	4555	20	**0.44**
Gentianales	**Gentianaceae**	1655	4	**0.24**
Gentianales	**Rubiaceae**	13150	4	**0.03**
Lamiales	**Acanthaceae**	4000	13	**0.33**
Lamiales	**Bignoniaceae**	800	8	**1.00**
Lamiales	**Lamiaceae**	7173	5	**0.07**
Lamiales	**Orobanchaceae**	2060	12	**0.58**
Lamiales	**Phrymaceae**	188	3	**1.60**
Lamiales	**Plantaginaceae**	1900	10	**0.53**
Lamiales	**Scrophulariaceae**	1800	10	**0.56**
Lamiales	**Verbenaceae**	918	12	**1.31**
Lamiales	**Boraginaceae**	2755	14	**0.51**
Laurales	**Lauraceae**	2500	2	**0.08**
Magnoliales	**Annonaceae**	2220	1	**0.05**
Malpighiales	**Calophyllaceae**	460	1	**0.22**
Malpighiales	**Chrysobalanaceae**	460	1	**0.22**
Malpighiales	**Clusiaceae**	595	1	**0.17**
Malpighiales	Elatinaceae	35	2	5.71

Order	Family	Number of Species	Number of Halophytes	Percentage of Halophytes
Malpighiales	**Euphorbiaceae**	5735	17	**0.30**
Malpighiales	**Hypericaceae**	560	1	**0.18**
Malpighiales	**Linaceae**	300	1	**0.33**
Malpighiales	**Phyllanthaceae**	1745	2	**0.11**
Malpighiales	**Putranjivaceae**	210	1	**0.48**
Malpighiales	**Rhizophoraceae**	149	19	**12.75**
Malpighiales	**Salicaceae**	1010	3	**0.30**
Malvales	**Malvaceae**	4225	27	**0.64**
Malvales	**Thymelaeaceae**	891	2	**0.22**
Myrtales	**Combretaceae**	500	10	**2.00**
Myrtales	**Lythraceae**	620	9	**1.45**
Myrtales	**Myrtaceae**	4620	22	**0.48**
Myrtales	**Onagraceae**	656	1	**0.15**
Nymphaeales	**Nymphaeaceae**	58	1	**1.72**
Pandanales	**Pandanaceae**	885	10	**1.13**
Picramniales	Picramniaceae	49	1	2.04
Piperales	**Piperaceae**	3615	1	**0.03**
Piperales	Saururaceae	6	1	16.67
Poales	**Bromeliaceae**	1770	2	**0.11**
Poales	**Cyperaceae**	5430	70	**1.29**
Poales	Flagellariaceae	4	1	25.00
Poales	**Juncaceae**	430	14	**3.26**
Poales	**Poaceae**	11160	212	**1.90**
Poales	**Restionaceae**	500	1	**0.20**
Poales	Typhaceae	25	6	24.00
Ranunculales	**Ranunculaceae**	2525	4	**0.16**
Rosales	Elaeagnaceae	45	2	4.44
Rosales	**Moraceae**	1125	5	**0.44**
Rosales	**Rhamnaceae**	925	5	**0.54**
Rosales	**Rosaceae**	2520	6	**0.24**
Rosales	Ulmaceae	35	1	2.86
Sapindales	**Anacardiaceae**	873	2	**0.23**
Sapindales	**Meliaceae**	615	3	**0.49**
Sapindales	Nitrariaceae	16	7	43.75
Sapindales	**Rutaceae**	2070	2	**0.10**
Sapindales	**Simaroubaceae**	110	1	**0.91**
Saxifragales	**Crassulaceae**	1370	1	**0.07**
Saxifragales	Cynomoriaceae	2	1	50.00
Solanales	**Convolvulaceae**	1625	14	**0.86**
Solanales	**Solanaceae**	2460	29	**1.18**
Vitales	**Vitaceae**	850	1	**0.12**
Zygophyllales	**Zygophyllaceae**	285	15	**5.26**

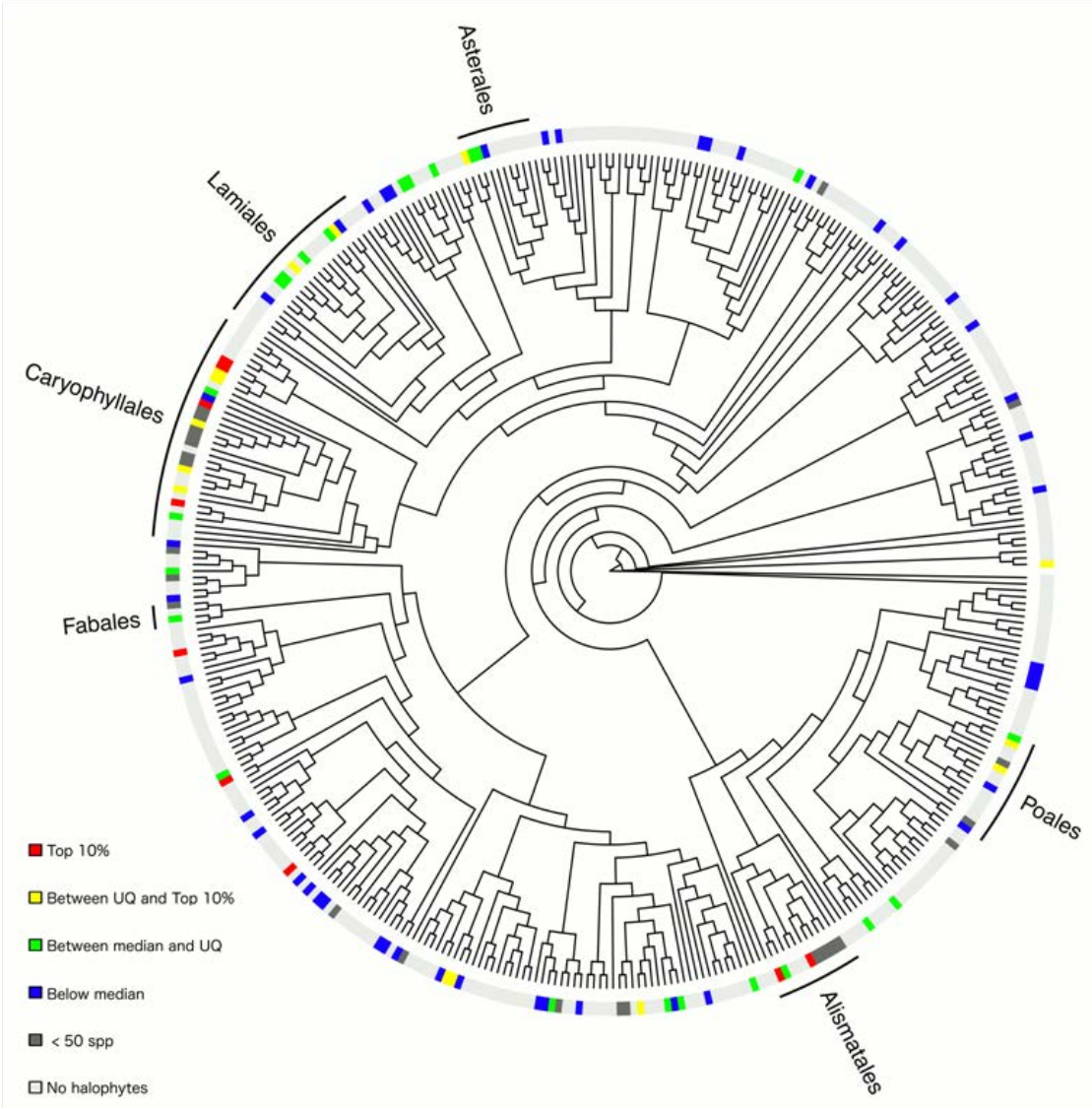


Figure 1: Phylogenetic tree of 401 APG III families extracted from a published angiosperm phylogeny (Smith et al., 2011). For each family, one representative taxon was selected based on maximum alignment length. Coloured tips represent families containing halophytes. Families were ranked by percentage of halophytes (see Table 1). Each tip is coloured based on the relative position of each family based on this ranking. We present families that were placed in top 10% of halophyte proportion (red), between the Upper Quartile (UQ) and the top 10% (yellow), between the median and the UQ (green), and below the median (blue). Families containing fewer than 50 species were not ranked and are shown in dark grey. Orders labelled on the phylogenetic tree contain at least 50 halophytes. The Figure was drawn with the R package "diversitree" (FitzJohn, 2012).

Thesis chapters

Chapter 2

The preliminary analysis of the phylogenetic patterns of salt tolerance in the angiosperms presented above suggests that salt tolerance has evolved a number of times among angiosperm families, but the evolutionary patterns of salt tolerance among angiosperm species is still largely unknown. In the Poaceae, it appears that salt tolerance has evolved over 70 times among species (Bennett et al. 2013). The inferred evolutionary origins of salt tolerance in the grasses appear close to the tips of the phylogeny, and many of the origins only give rise to one or a few halophytes. This pattern is intriguing considering that salt tolerance involves many physiological or anatomical changes and has been incredibly difficulty to breed in crops and model organisms (Flowers and Yeo 1995; Colmer et al. 2005; Ashraf and Akram 2009; Ashraf and Foolad 2012). However, it is unclear whether the repeated evolution of salt tolerance is unique to the evolutionary history of the grasses, or whether it represents a general feature of salt tolerance evolution in the angiosperms.

To address this question, in Chapter 2, I examine the phylogenetic distribution of halophytes across angiosperms. I first expand the halophyte list used in the preliminary above by adding published lists of species identified as halophytes from different geographic regions. Using this expanded list, I establish whether the distribution of halophytes is non-random with respect to angiosperm diversity and also identify families with significantly more or fewer halophytes than predicted by both the proportion of known halophytes and the species diversity of each family. I then explore the phylogenetic distribution of halophytes in a set of angiosperm families using several phylogenetic comparative analyses to establish whether the pattern found in the grasses is common among angiosperms.

Chapter 3

One explanation for why salt tolerance has evolved many times in angiosperms is that it builds on a suite of stress tolerance traits that can equip lineages to adapt to many different harsh environmental conditions. One way to test this idea is to compare the

phylogenetic relationship between halophytes and species with other traits associated with adaptation to abiotic stress. For example, a few recent studies have found evidence of an evolutionary relationship between salt tolerance and C_4 photosynthesis (associated with aridity tolerance, Bromham and Bennett 2014) and salt and alkalinity tolerance (Bui et al 2013; Saslis-Lagoudakis et al. 2015). Although relatively unexplored, these studies can provide a backdrop for more fine scale studies on the evolution of physiological mechanisms that support abiotic stress tolerance.

Several studies suggest that salt tolerance may be linked to another rare plant trait associated with harsh environments - the ability to tolerate and accumulate heavy metals (Ghnaya et al. 2007; Manousaki and Kalogerakis 2011; Rozema and Schat 2013). Although heavy metals are toxic to most plants, there is a rare group of plant species, heavy metal hyperaccumulators, that can not only tolerate but also take up high concentrations of heavy metals into their tissues (Baker and Brooks 1989). Several observations suggest an association between salt tolerance and heavy metal hyperaccumulation. First, both salt and heavy metals induce osmotic and oxidative stress, and halophytes and hyperaccumulators use some of the same mechanisms to combat these stresses. For example, both halophytes (Glenn et al. 1999; Blumwald 2000; Munns and Tester 2008) and heavy metal hyperaccumulators (Schat et al. 1997; Sharma and Dietz 2006; Lefèvre et al. 2009) commonly produce osmoprotectants (compatible solutes) to maintain osmotic potential and mitigate the damaging effects of excess salinity in cells. In some cases, halophytes and heavy metal hyperaccumulators produce the same compatible solutes (e.g., proline, Stewart and Lee 1974; Flowers et al. 1977; Schat et al. 1997; Sharma and Dietz 2006). Second, several studies have identified halophytes that can also accumulate heavy metals (Jordan et al. 2002; Kadukova et al. 2008; Redondo-Gómez et al. 2010; Redondo-Gómez 2013). Third, phylogenetic and taxonomic evidence suggests that salt tolerance (Flowers et al. 2010; Bennett et al. 2013; Saslis-Lagoudakis et al. 2014) and heavy metal hyperaccumulation (Cappa and Pilon-Smits 2014) have both evolved many times in the angiosperms, and in many of the same taxonomic groups (e.g., Brassicaceae, Rascio and Navari-Izzo 2011 and Asteraceae, Vara Prasad and de Oliveira Freitas 2003). Given the physiological, taxonomic, and evolutionary similarities between salt tolerance and heavy metal hyperaccumulation, can we find evidence of an evolutionary association between these ecophysiological strategies?

To address this question, in Chapter 3, I formally test the generality of the relationship between these two abilities across a wide range of angiosperm groups. I use taxonomic analyses to test whether halophytes and hyperaccumulators are significantly associated among angiosperm families and whether there are more species identified as both a halophyte and heavy metal hyperaccumulator than expected given the rarity of both abilities. I also use a phylogenetic analysis to test whether on average halophytes and heavy metal hyperaccumulators are more closely related than expected if the two abilities had evolved independently.

Chapter 4

One reason that salt tolerance has been difficult to breed could be that salt tolerance is a costly ability to develop and maintain (Flowers and Yeo 1995; Yeo1983). In addition to the cost of adjusting osmotic potential, expending energy on ion regulation, and producing compatible solutes, halophytes may also experience oxidative stress from the toxic effect of salinity on plant tissues, which could lead to increased mutation rates. We can expect that genes associated with salinity tolerance, for example ones that are important for preventing oxidative damage (Dennis and Shimmin 1997), are likely to be selected and to change in halophytes compared to their non-salt tolerant relatives. However, if oxidative damage causes an increase in the generation of mutations genome-wide, or causes mutations in regions that effect the replication and repair of DNA, salinity might lead to an overall increase in the rate at which nucleotide changes become ubiquitous among halophyte populations compared to their non-salt tolerant relatives. Few studies have examined the influence of environmental factors on rates of molecular evolution (e.g., Hebert et al. 2002; Whittle 2006; Groussin and Gouy 2011; Gillman and Wright 2013), but some suggest that salinity may increase rates of molecular evolution, the pace at which DNA accumulates changes in nucleotides, in a diverse set of organisms (Dennis and Shimmin 1997; Hebert et al. 2002; Wägele et al. 2003; Baxevanis et al. 2006; Logares et al. 2010). However, some studies have found no difference in rates of molecular evolution between salt tolerant species and their close, non-salt tolerant relatives (Whittle 2006; Logares et al. 2009). Thus far these studies have examined relatively few genes and many have only examined genes in one genome, so it is still unknown whether salinity has a consistent effect on rates of molecular evolution. In Chapter 4, I compare DNA sequences from multiple genes in the chloroplast, mitochondrial, and nuclear genomes to test whether halophytes have

consistently increased rates of molecular evolution compared to their non-salt tolerant relatives.

Chapter 5

Although the majority of crop species are not salt tolerant, several studies have found that landraces and wild relatives of several crops contain haplotypes that are tolerant to many abiotic stresses, including salinity (e.g., soybean, Guan et al. 2014, barley, Colmer et al. 2005). Many crop genepools have lost genetic variability (Abbo et al. 2003), which could explain why some crops lack salt tolerance even though they have salt tolerant wild relatives (e.g., barley relative *Hordeum maritimum*, Colmer et al. 2005 and sugar beet relative *Beta maritima*, Rozema et al. 2014) and why some crops have lost salt tolerance during domestication (e.g., soybean, Hyten et al. 2006). Without key ancestral traits, however, it has been suggested that traditional breeding may never lead to salt tolerant varieties as the crop genepools have lost the genetic variation necessary to develop salt tolerance (Colmer et al. 2005). Domestication involves the systematic reduction of population size (e.g., bottlenecks), which over time can remove some haplotypes from the crop genepool. But there is also evidence that domestication may more broadly effect the rate of molecular evolution (Lu et al. 2006; Purugganan and Fuller 2011), which could contribute to how domesticated populations can adapt or generate novel variation over evolutionary short time scales. The current evidence is conflicting, suggesting that in some plant and animal lineages domestication may increase rates (Gu et al. 2005; Lu et al. 2006; Cruz et al. 2008; Hughes 2013), while in others they may not (Rokas 2009; Purugganan and Fuller 2011). In Chapter 5, I explore whether there is a consistent effect of domestication on rates of molecular evolution.

This project was conducted during the first year of my PhD with my co-supervisor Robert Lanfear while my primary supervisor Lindell Bromham was on leave. So although breeding for salt tolerance in crops is a related issue, I explored rates of molecular evolution in domesticated lineages of mammals and birds. Several studies have reported increased rates of molecular evolution in domesticated animals compared to their wild relatives (Björnerfeldt et al. 2006; Cruz et al. 2008; Wang et al. 2011; Hughes 2013), but relatively few lineages have been examined. These studies suggest that domestication consistently increases rates of molecular evolution because of relaxed selective constraint – changing the lifestyle and thus metabolic demands of

domesticated individuals, allowing for the accumulation of slightly deleterious mutations in the genome. To test whether domestication consistently increases rates of molecular evolution, I used a comparative phylogenetic approach to compare rates of molecular evolution in the complete or nearly complete mitochondrial genomes of 17 pairs of domesticated animals and their closely related wild relatives. Although this project does not directly relate to salt tolerance, this was an introduction into learning phylogenetic comparative methods and greatly contributed to the analytical skills used in my other chapters (particularly Chapter 4). Domestication is also an excellent example of how lineages can evolve and accumulate detectable changes over evolutionary short time scales, which may be related to how salt tolerance has evolved repeatedly among closely related species (Bennett et al. 2013) and to the variation in salt tolerance among closely related species (Greenwood and MacFarlane 2009).

References

Abbo, S., J. Berger, and N. C. Turner. 2003. Viewpoint: evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. Functional Plant Biology 30:1081.

Albert, R. 1975. Salt regulation in halophytes. Oecologia 21:57–71.

Aronson, J. A. 1989. HALOPH: a data base of salt tolerant plants of the world. Office of Arid Lands Studies, University of Arizona, Tucson, AZ.

Ashraf, M., and M. R. Foolad. 2012. Crop breeding for salt tolerance in the era of molecular markers and marker-assisted selection. Plant Breeding 132:10-20.

Ashraf, M., and N. A. Akram. 2009. Improving salinity tolerance of plants through conventional breeding and genetic engineering: an analytical comparison. Biotechnology advances 27:744–752.

Baker, A. J. M., and R. R. Brooks. 1989. Terrestrial higher plants which hyper-accumulate metallic elements - a review of their distribution, ecology and phytochemistry. Biorecovery 1:81–126.

Ball, M. C. 1988. Ecophysiology of mangroves. Trees 2:129–142.

Ball, M. C. 1998. Mangrove species richness in relation to salinity and waterlogging: a case study along the Adelaide river floodplain, Northern Australia. Global Ecology and Biogeography Letters 7:73–82.

Baxevanis, A. D., I. Kappas, and T. J. Abatzopoulos. 2006. Molecular phylogenetics and asexuality in the brine shrimp *Artemia*. Molecular Phylogenetics and Evolution 40:724–738.

Bennett, S. J., E. G. Barrett-Lennard, and T. D. Colmer. 2009. Salinity and waterlogging as constraints to saltland pasture production: a review. Agriculture, Ecosystems and Environment 129:349–360.

Bennett, T. H., T. J. Flowers, and L. Bromham. 2013. Repeated evolution of salt-tolerance in grasses. Biology letters 9:20130029–20130029.

Björnerfeldt, S., M. T. Webster, and C. Vilà. 2006. Relaxation of selective constraint on dog mitochondrial DNA following domestication. Genome Research 16:990–994.

Blumwald, E. 2000. Sodium transport and salt tolerance in plants. Current Opinion in Cell Biology 12:431–434.

BOSTID. 1990. Saline agriculture: salt-tolerant plants for developing countries. National Academies Press, Washington D.C.

Boyle, B., N. Hopkins, Z. Lu, J. A. R. Garay, D. Mozzherin, T. Rees, N. Matasci, M. L. Narro, W. H. Piel, S. J. Mckay, S. Lowry, C. Freeland, R. K. Peet, and B. J. Enquist. 2013. The taxonomic name resolution service: an online tool for automated standardization of plant names. BMC Bioinformatics 14:1–1.

Bromham, L., and T. H. Bennett. 2014. Salt tolerance evolves more frequently in C_4 grass lineages. Journal of Evolutionary Biology 27:653–659.

Bromham, L. 2015. Macroevolutionary patterns of salt tolerance in angiosperms. Annals of Botany 115:333–341.

Bui, E. N., A. Thornhill, and J. T. Miller. 2014. Salt- and alkaline-tolerance are linked in *Acacia*. Biology letters 10:20140278–20140278.

Cao, D., C. C. Baskin, J. M. Baskin, F. Yang, and Z. Huang. 2014. Dormancy cycling and persistence of seeds in soil of a cold desert halophyte shrub. Annals of Botany 113:171–179.

Cappa, J. J., and E. A. H. Pilon-Smits. 2014. Evolutionary aspects of elemental hyperaccumulation. Planta 239:267–275.

Cayuela, L., Í. Granzow-de la Cerda, F. S. Albuquerque, and D. J. Golicher. 2012. Taxonstand: An R package for species names standardisation in vegetation databases. Methods in Ecology and Evolution 3:1078–1083.

Cheeseman, J. M. 1988. Mechanisms of salinity tolerance in plants. Plant Physiology 87:547–550.

Cheeseman, J. M. 2014. The evolution of halophytes, glycophytes and crops, and its implications for food security under saline conditions. New Phytologist 206:557-570.

Colmer, T. D., and T. J. Flowers. 2008. Flooding tolerance in halophytes. New Phytologist 179:964–974.

Colmer, T. D., R. Munns, and T. J. Flowers. 2005. Improving salt tolerance of wheat and barley: future prospects. Australian Journal of Experimental Agriculture 45:1425.

Cruz, F., C. Vilà, and M. T. Webster. 2008. The legacy of domestication: accumulation of deleterious mutations in the dog genome. Molecular Biology and Evolution 25:2331–2336.

Dennis, P. P., and L. C. Shimmin. 1997. Evolutionary divergence and salinity-mediated selection in halophilic archaea. Microbiology and Molecular Biology Reviews 61:90–104.

Djanaguiraman, M., Sheeba, J. A., Shanker, A. K., Devi, D. D., and U. Bangarusamy. 2006. Rice can acclimate to lethal level of salinity by pretreatment with sublethal level of salinity through osmotic adjustment. Plant and Soil 284:363–373.

FitzJohn, R. G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods in Ecology and Evolution 3:1084–1092.

Flowers, T. J., and A. R. Yeo. 1995. Breeding for salinity resistance in crop plants: where next? Australian Journal of Plant Physiology 22:875–884.

Flowers, T. J., and A. R. Yeo. 1986. Ion relations of plants under drought and salinity. Australian Journal of Plant Physiology 13:75–91.

Flowers, T. J., and T. D. Colmer. 2008. Salinity tolerance in halophytes. New

Phytologist 179:945–963.

Flowers, T. J., H. K. Galal, and L. Bromham. 2010. Evolution of halophytes: multiple origins of salt tolerance in land plants. Functional Plant Biology 37:604.

Flowers, T., P. F. Troke, and A. R. Yeo. 1977. The mechanism of salt tolerance in halophytes. Annual Review of Plant Physiology 28:89–121.

Ghnaya, T., I. Slama, D. Messedi, C. Grignon, M. H. Ghorbel, and C. Abdelly. 2007. Effects of Cd^{2+} on K^{+}, Ca^{2+} and N uptake in two halophytes *Sesuvium portulacastrum* and *Mesembryanthemum crystallinum*: Consequences on growth. Chemosphere 67:72–79.

Gillman, L. N., and S. D. Wright. 2013. Species richness and evolutionary speed: the influence of temperature, water and area. Journal of Biogeography 41:39–51.

Glenn, E. P., J. J. Brown, and E. Blumwald. 1999. Salt tolerance and crop potential of halophytes. Critical Reviews in Plant Science 18:227–255.

Gómez-Pando, L. R., R. Álvarez-Castro, and A. Eguiluz-de la Barra. 2010. Effect of salt stress on Peruvian germplasm of *Chenopodium quinoa* Willd.: a promising crop. Journal of Agronomy and Crop Science 196:391–396.

Greenway, H., and R. Munns. 1980. Mechanisms of salt tolerance in nonhalophytes. Annual Review of Plant Physiology 31:149–190.

Greenwood, M. E., and G. R. MacFarlane. 2009. Effects of salinity on competitive interactions between two *Juncus* species. Aquatic Botany 90:23–29.

Groussin, M., and M. Gouy. 2011. Adaptation to environmental temperature is a major determinant of molecular evolutionary rates in Archaea. Molecular Biology and Evolution 28:2661–2674.

Gu, Z., L. David, D. Petrov, T. Jones, R. W. Davis, and L. M. Steinmetz. 2005. Elevated evolutionary rates in the laboratory strain of *Saccharomyces cerevisiae*. Proceedings of the National Academy of Sciences of the United States of America 102:1092–1097.

Guan, R., Y. Qu, Y. Guo, L. Yu, Y. Liu, J. Jiang, J. Chen, Y. Ren, G. Liu, L. Tian, L. Jin, Z. Liu, H. Hong, R. Chang, M. Gilliham, and L. Qiu. 2014. Salinity tolerance in soybean is modulated by natural variation in GmSALT3. The Plant Journal 80:937–950.

Guvensen, A., G. Gork, & M. Özturk. 2006. An overview of the halophytes in Turkey. In M. A. Khan, G. S. Kust, H.-J. Barth, & B. Böer (Eds.), Sabkha ecosystems. Vol: II. Springer, Netherlands, pp. 9–30.

Hebert, P. D. N., E. A. Remigio, J. K. Colbourne, D. J. Taylor, and C. C. Wilson. 2002. Accelerated molecular evolution in halophilic crustaceans. Evolution 56:909–926.

Hughes, A. L. 2013. Accumulation of slightly deleterious mutations in the mitochondrial genome: a hallmark of animal domestication. Gene 515:28–33.

Hyten, D. L., Q. Song, Y. Zhu, I.-Y. Choi, R. L. Nelson, J. M. Costa, J. E. Specht, R. C. Shoemaker, and P. B. Cregan. 2006. Impacts of genetic bottlenecks on soybean genome

diversity. Proceedings of the National Academy of Sciences of the United States of America 103:16666–16671.

Jaradat, A. A. 2003. Halophytes for sustainable biosaline farming systems in the Middle East. In: A. S. Alsharhan, W. W. Wood, A. S. Goudie, A. Fowler, and E. M. Abdellatif, (Eds.) Desertification in the third millennium. A A Balkema, Rotterdam, The Netherlands, pp. 1–18.

Jordan, F., M. Robin Abbott, R. M. Maier, and E. P. Glenn. 2002. A comparison of chelator-facilitated metal uptake by a halophyte and a glycophyte. Environmental toxicology and chemistry 21:2698–2704.

Kadereit, G., D. Ackerly, and M. D. Pirie. 2012. A broader model for C_4 photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae s.s.). Proceedings of the Royal Society B: Biological Sciences 279:3304–3311.

Kadukova, J., E. Manousaki, and N. Kalogerakis. 2008. Pb and Cd Accumulation and Phyto-Excretion by Salt Cedar (*Tamarix smyrnensis* Bunge). International Journal of Phytoremediation 10:31–46.

Lefèvre, I., G. Marchal, P. Meerts, E. Corréal, and S. Lutts. 2009. Chloride salinity reduces cadmium accumulation by the Mediterranean halophyte species *Atriplex halimus* L. Environmental and Experimental Botany 65:142–152.

Les, D., M. Cleland, and M. Waycott. 1997. Phylogenetic studies in Alismatidae, II: Evolution of marine angiosperms (seagrasses) and hydrophily. Systematic Botany 22:443–463.

Logares, R., J. Brate, F. Heinrich, K. Shalchian-Tabrizi, and S. Bertilsson. 2010. Infrequent transitions between saline and fresh waters in one of the most abundant microbial lineages (SAR11). Molecular Biology and Evolution 27:347–357.

Logares, R., J. Bråte, S. Bertilsson, J. L. Clasen, K. Shalchian-Tabrizi, and K. Rengefors. 2009. Infrequent marine–freshwater transitions in the microbial world. Trends in Microbiology 17:414–422.

Lu, J., T. Tang, H. Tang, J. Huang, S. Shi, and C.-I. Wu. 2006. The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. Trends in Genetics 22:126–131.

Maddison, W. P., and D. R. Maddison. 2006. Mesquite: a modular system for evolutionary analysis. Available at http://mesquiteproject.org.

Manousaki, E., and N. Kalogerakis. 2011. Halophytes—an emerging trend in phytoremediation. International Journal of Phytoremediation 13:959–969.

Menzel, U., and H. Lieth. 2003. Halophyte database version 2.0. In: Cash crop halophytes: recent studies. Kluwer Academic Publishers, Dordrecht, pp. 221–250.

Munns, R. 2002. Comparative physiology of salt and water stress. Plant, Cell and Environment 25:239–250.

Munns, R. 2005. Genes and salt tolerance: bringing them together. New Phytologist 167:645–663.

Munns, R. 2011. Plant adaptations to salt and water stress: differences and commonalities. Advances in Botanical Research57:1–32.

Munns, R., and M. Tester. 2008. Mechanisms of salinity tolerance. Annual Review of Plant Biology 59:651–681.

Munns, R., R. A. James, and A. Läuchli. 2006. Approaches to increasing the salt tolerance of wheat and other cereals. Journal of Experimental Botany 57:1025–1043.

Öztürk, M., A. Guvensen, S. Sakçali, & G. Gork. 2008. Halophyte plant diversity in the Irano-Turanian phytogeographical region of Turkey. In C. Abdelly, M. Öztürk, M. Ashraf, & C. Grignon (Eds.), Biosaline Agriculture and High Salinity Tolerance. Birkhäuser, Basel, pp. 141–155.

Pandolfi, C., Mancuso, S., and S. Shabala. 2012. Physiology of acclimation to salinity stress in pea (*Pisum sativum*). Environmental and Experimental Botany 84:44–51.

Panta, S., T. Flowers, P. Lane, R. Doyle, G. Haros, and S. Shabala. 2014. Halophyte agriculture: Success stories. Environmental and Experimental Botany 107:71–83.

Pennings, S. C., and R. M. Callaway. 1992. Salt marsh plant zonation - the relative importance of competition and physical factors. Ecology 73:681–690.

Purugganan, M. D., and D. Q. Fuller. 2011. Archaeological data reveal slow rates of evolution during plant domestication. Evolution 65:171–183.

R Core Team. 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rascio, N., and F. Navari-Izzo. 2011. Heavy metal hyperaccumulating plants: how and why do they do it? And what makes them so interesting? Plant Science 180:169–181.

Redondo-Gómez, S. 2013. Bioaccumulation of heavy metals in *Spartina*. Functional Plant Biology 40:913–921.

Redondo-Gómez, S., E. Mateos-Naranjo, and L. Andrades-Moreno. 2010. Accumulation and tolerance characteristics of cadmium in a halophytic Cd-hyperaccumulator, *Arthrocnemum macrostachyum*. Journal of Hazardous Materials 184:299–307.

Rengasamy, P. 2006. World salinization with emphasis on Australia. Journal of Experimental Botany 57:1017–1023.

Rokas, A. 2009. The effect of domestication on the fungal proteome. Trends in Genetics 25:60–63.

Rozema, J., and H. Schat. 2013. Salt tolerance of halophytes, research questions reviewed in the perspective of saline agriculture. Environmental and Experimental Botany 92:83–95.

Rozema, J., and T. Flowers. 2008. Crops for a salinized world. Science 322:1478–1480.

Rozema, J., D. Cornelisse, Y. Zhang, H. Li, B. Bruning, D. Katschnig, R. Broekman, B. Ji, and P. van Bodegom. 2014. Comparing salt tolerance of beet cultivars and their

halophytic ancestor: consequences of domestication and breeding programs. AoB plants:plu083.

Ruan, C.-J., J. A. T. da Silva, S. Mopper, P. Qin, and S. Lutts. 2010. Halophyte improvement for a salinized world. Critical Reviews in Plant Science 29:329–359.

Saslis-Lagoudakis, C. H., X. Hua, E. Bui, C. Moray, and L. Bromham. 2015. Predicting species' tolerance to salinity and alkalinity using distribution data and geochemical modelling: a case study using Australian grasses. Annals of Botany 115:343–351.

Saslis-Lagoudakis, C. H., Moray, C., and Bromham, L. 2014. Evolution of salt tolerance in angiosperms: a phylogenetic approach. In: N. Rajakaruna, R. S. Boyd, & T. B. Harris (Eds.) Plant ecology and evolution in harsh environments. Nova Science Publishers, Hauppauge, NY, pp. 77–95.

Schat, H., S. S. Sharma, and R. Vooijs. 1997. Heavy metal-induced accumulation of free proline in a metal-tolerant and a nontolerant ecotype of *Silene vulgaris*. Physiologia Plantarum 101:477–482.

Schofield, R. V., and M. J. Kirkby. 2003. Application of salinization indicators and initial development of potential global soil salinization scenario under climatic change. Global Biogeochemical Cycles 17:1078.

Shannon, M. C. 1997. Adaptation of plants to salinity. Advances in Agronomy 60:75–120.

Shannon, M. C. 1985. Principles and strategies in breeding for higher salt tolerance. Plant Soil 89:227–241.

Sharma, S. S., and K.-J. Dietz. 2006. The significance of amino acids and amino acid-derived molecules in plant responses and adaptation to heavy metal stress. Journal of Experimental Botany 57:711–726.

Smith, S. A., J. M. Beaulieu, A. Stamatakis, and M. J. Donoghue. 2011. Understanding angiosperm diversification using small and large phylogenetic trees. American Journal of Botany 98:404–414.

Stewart, G. R., and J. A. Lee. 1974. The role of proline accumulation in halophytes. Planta 120:279–289.

Tester, M., and P. Langridge. 2010. Breeding technologies to increase crop production in a changing world. Science 327:818.

The Plant List 2010. The Plant List version 2. http://www.theplantlist.org/.

Vara Prasad, M. N., and H. M. de Oliveira Freitas. 2003. Metal hyperaccumulation in plants: biodiversity prospecting for phytoremediation technology. Electronic Journal of Biotechnology 6:285–321.

Ventura, Y., and M. Sagi. 2013. Halophyte crop cultivation: The case for *Salicornia* and *Sarcocornia*. Environmental and Experimental Botany 92:144–153.

Ventura, Y. Eshel, A., Pasternak, D., and M. Sagi. 2015. The development of halophyte-based agriculture: past and present. Annals of Botany 115:529–540.

Waisel, Y. 1972. The biology of halophytes. Academic Press, London.

Wang, W., B. Vinocur, and A. Altman. 2003. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta 218:1–14.

Wang, Z., T. Yonezawa, B. Liu, T. Ma, X. Shen, J. Su, S. Guo, M. Hasegawa, and J. Liu. 2011. Domestication relaxed selective constraints on the yak mitochondrial genome. Molecular Biology and Evolution 28:1553–1556.

Wägele, J.-W., B. Holland, H. Dreyer, and B. Hackethal. 2003. Searching factors causing implausible non-monophyly: ssu rDNA phylogeny of *Isopoda asellota* (Crustacea: Peracarida) and faster evolution in marine than in freshwater habitats. Molecular Phylogenetics and Evolution 28:536–551.

Whittle, C. A. 2006. The influence of environmental factors, the pollen : ovule ratio and seed bank persistence on molecular evolutionary rates in plants. Journal of Evolutionary Biology 19:302–308.

Witcombe, J. R., P. A. Hollington, C. J. Howarth, S. Reader, and K. A. Steele. 2008. Breeding for abiotic stresses for sustainable agriculture. Philosophical Transactions of the Royal Society of London B: Biological Sciences 363:703–716.

Yeo, A. R. 1983. Salinity resistance: Physiologies and prices. Physiologia Plantarum 58:214–222.

Zhao, K., Song, J., Feng, G., Zhao, M., and J. Liu. 2011. Species, types, distribution, and economic potential of halophytes in China. Plant and Soil 342:495–509.

Chapter 2

Salt tolerance is evolutionarily labile in a diverse set of angiosperm families

Camile Moray, Xia Hua, Lindell Bromham

Published: BMC Evolutionary Biology. 2015. 15:90.

Abstract

Background: Salt tolerance in plants is rare, yet it is found across a diverse set of taxonomic groups. This suggests that, although salt tolerance often involves a set of complex traits, it has evolved many times independently in different angiosperm lineages. However, the pattern of evolution of salt tolerance can vary dramatically between families. A recent phylogenetic study of the Chenopodiaceae (goosefoot family) concluded that salt tolerance has a conserved evolutionary pattern, being gained early in the evolution of the lineage then retained by most species in the family. Conversely, a phylogenetic study of the Poaceae (grass family) suggested over 70 independent gains of salt tolerance, most giving rise to only one or a few salt tolerant species. Here, we use a phylogenetic approach to explore the macroevolutionary patterns of salt tolerance in a sample of angiosperm families, in order to ask whether either of these two patterns – deep and conserved or shallow and labile - represents a common mode of salt tolerance evolution. We analyze the distribution of halophyte species across the angiosperms and identify families with more or less halophytes than expected under a random model. Then, we explore the phylogenetic distribution of halophytes in 22 families using phylogenetic comparative methods.

Results: We find that salt tolerance species have been reported from over one-third of angiosperm families, but that salt tolerant species are not distributed evenly across angiosperm families. We find that salt tolerance has been gained hundreds of times over the history of the angiosperms. In a few families, we find deep and conserved gains of salt tolerance, but in the majority of families analyzed, we find that the pattern of salt tolerant species is best explained by multiple independent gains that occur near the tips of the phylogeny and often give rise to only one or a few halophytes.

Conclusions: Our results suggest that the pattern of many independent gains of salt tolerance near the tips of the phylogeny is found in many angiosperm families. This suggests that the pattern reported in the grasses of high evolutionary lability may be a common feature of salt tolerance evolution in angiosperms.

Keywords: angiosperms, comparative method, halophyte, macroevolution, repeated evolution

Background

Only 1 - 2% of angiosperm species are known to be halophytes, able to live and reproduce in saline soils [1, 2]. The rarity of salt tolerance is unsurprising considering it is a costly and complex ecological strategy; halophytes may have modifications to many parts of their physiology and anatomy in order to combat the damaging effects of osmotic and metabolic stress, which can cause impaired growth and reproduction [2-4]. However, halophytes are found in a wide range of angiosperm families and they occupy diverse habitats worldwide.

Salt tolerance has also clearly evolved multiple times in angiosperms [5, 6]. The evolutionary patterns of salt tolerance in plants have been studied in detail in only a few taxonomic groups, and these studies have revealed two very different patterns of salt tolerance evolution. In one well-studied group, the chenopods (Chenopodiaceae), salt tolerance appears to be phylogenetically conserved [7], arising only once or twice in the history of the group, then being retained in a large proportion of species in the family. In contrast, a study on the grass family (Poaceae) estimated that there have been at least 70 origins of salt tolerance within the family [8]. Most of these inferred origins were near the terminal taxa (tips) of the phylogeny, suggesting multiple shallow origins, each giving rise to only one or a few salt tolerant species. The pattern of salt tolerance evolution inferred in the Poaceae is interesting because it suggests that, at least in the grasses, salt tolerance has evolved repeatedly in a range of lineages, despite the complexity of salt tolerance adaptations. However, the observation that salt tolerance does not persist over long evolutionary timescales in the grasses may indicate that while salt tolerance is easy to gain, it is also frequently lost through trait reversal or extinction, implying that there are costs associated with the adoption of salt tolerance.

These two different phylogenetic patterns suggest very different macroevolutionary dynamics. Salt tolerance is highly conserved in the chenopods, with a large number of salt tolerant species arising from only a few independent origins. But in the grasses, salt tolerance is highly labile, in the sense that it is gained and lost relatively frequently. Which, if either, of these patterns is observed in other families of angiosperms? To answer this question, we use a phylogenetic comparative approach to investigate and characterize patterns of halophyte diversity and evolution among angiosperm families.

Halophytes use a variety of physiological and anatomical traits to survive in saline habitats, and these traits can vary between species. Some halophytes exhibit complex anatomical modifications like salt glands or hairs, but most halophytes rely on osmotic regulation, modifying existing physiological mechanisms to mitigate salinity levels within the plant [9, 10]. These strategies can also vary amongst closely related halophytes and among halophytes that occupy similar habitats, for example the differential presence of succulence among closely related chenopods [7] or salt glands among phylogenetically diverse mangrove species [11]. Instead of identifying specific environmental or physiological differences between halophytes, we focus on the broad distribution of salt tolerance as an ecological strategy amongst angiosperms, at the family and species levels. We first examine how halophytes are distributed among the angiosperm families, identifying any families that have more or less halophytes than expected by chance. Then we use a number of phylogenetic measures to analyze the observed evolutionary patterns of salt tolerance in a sample of 22 angiosperm families. This sample includes large families with many known halophytes, including families with both more and less halophytes than expected.

Results

Halophyte Diversity

We found that the observed distribution of halophytes across angiosperm families was significantly nonrandom ($p < 0.001$). Of the 411 families included in the taxonomic analysis, 146 families have one or more known halophytes (See Methods and Additional files). We found that 51 of the 411 families have significantly more halophytes than expected by chance; examples include Amaranthaceae, Poaceae and Rhizophoraceae (see Table S1). 68 families have significantly fewer halophytes than expected by chance, for example Acanthaceae, Lamiaceae and Fabaceae.

Evolutionary Patterns

For each family analyzed, we created a family subtree that included all tips from a large angiosperm phylogeny [12] belonging to each family according to GenBank taxonomy (See Methods). In the 22 family subtrees analyzed, we observed a range of evolutionary patterns of salt tolerance (see Figure 1 and Figure S1). In general, evolutionary gains of salt tolerance appeared close to the tips, across the family subtrees. One measure used to assess the evolutionary patterns of salt tolerance across families was the number of tips

per origin (NoTO), the average number of taxa descending from each inferred evolutionary origin of salt tolerance in a family. The median value of NoTO across all 22 families analyzed was 1.3 and nineteen of the family subtrees had a NoTO value less than two. This observation indicates that the inferred gains of salt tolerance in these family subtrees typically give rise to less than two descendant halophyte tips. In contrast, a few families (Rhizophoraceae, Amaranthaceae and Tamaricaceae) had higher NoTO values than the other families, meaning that each gain of salt tolerance in these families is deeper in the subtrees and leads to comparatively larger clades of halophytes than observed in the other families analyzed. Tamaricaceae was the only family in our sample with significantly fewer salt tolerance gains than expected given the number of known halophytes and halophytic taxa were significantly clustered.

Over half of the families analyzed had a similar phylogenetic distribution to the pattern found in the grasses [8]. In these families, given the observed number of halophytes in each of the family subtrees, either 1) salt tolerance has evolved more times than expected under a Brownian motion model (significantly lower NoTO) and/or 2) clades of halophytes are less clustered than expected under Brownian motion model of trait evolution (significantly higher sum of sister clade differences (SSCD), see Methods) (Table 1). When comparing with the results from the angiosperm diversity analysis, we found that this labile evolutionary pattern is found in families with varying proportions of halophytes, including those with more, fewer, and within the expected number of halophytes based on family size.

Correlation tests suggest that there is no significant association between taxon sampling proportion and estimates of SSCD ($p = 0.660$, $\tau = 0.071$) or NoTO ($p = 0.728$, $\tau = 0.056$) estimates. The proportion of halophytic taxa in the subtrees is not correlated with SSCD ($p = 0.158$, $\tau = 0.230$) or NoTO p-values ($p = 0.087$, $\tau = 0.270$).

Discussion

Using a list of known halophytes assembled from a range of published sources, we find that one-third of angiosperm families contain species reported as being able to live in saline conditions. We show that the distribution of salt tolerant species among angiosperm families is not consistent with a random distribution, and that some families have significantly more halophytes than expected given the family size, while others have significantly fewer salt tolerant species.

Not only does the proportion of halophytes differ between families, but the phylogenetic distribution of salt tolerance within families also varies. Specifically, we set out to test whether the pattern of salt tolerance in grasses (Poaceae) – with many, shallow, scattered gains – was also found in other families. A few families show the opposite pattern, where salt tolerance has been gained deep in the family and retained by a large proportion of descendants [7]; examples include Tamaricaceae which contains the highly salt tolerant salt cedars and many species with specialized anatomical traits like salt glands [13]; Amaranthaceae, which includes the halophyte-rich groups formerly classified under Chenopodiaceae; and Rhizophoraceae, which contains many mangrove species. However, over half of the families analyzed show a pattern like the grasses, consistent with many, shallow gains of salt tolerance. The fact that we find this labile pattern of salt tolerance evolution in a phylogenetically diverse set of families with different proportions of halophytes suggests that the observed pattern of many independent gains of salt tolerance is not simply explained by the proportion of halophytes in a group.

One limitation of broad comparative analyses like this one is that we can only gain information from data on known halophytes and sequenced angiosperm taxa. Specifically, the limitations in data used in this study come from two main sources. One problem is the incidence of false negatives in the halophyte list. Most published lists of halophytes are based on observational data, and there are likely to be other salt tolerant species that have not been described in the literature or included in published lists. For example, there are likely many species living in non-saline habitats that have the capacity for salt tolerance but have not yet been formally tested. One solution to improve future analyses is to move away from list-based methods drawn from single species experiments and observational data. For example, it may be possible for phylogenetic studies to contribute to the identification of salt tolerance lineages, for example by using use phyloprediction [14] or geochemical modeling to identify lineages that are likely to be salt tolerant [15, 16].

A second potential source of error is incomplete phylogenetic sampling. The phylogeny used in this study includes about 20% of known angiosperm species, so there are some known salt tolerant taxa that are not included in this tree (see Table 1 for details). Correlation tests did not indicate any consistent effect of the proportion of total species sampled in a family or of the proportion of halophytes in the subtrees on the results of

NoTO and SSCD, suggesting that sampling proportion does not significantly influence the results of our analysis. Increased sampling is unlikely to change the overarching pattern because salt tolerant taxa in most family subtrees with significant NoTO and SSCD are sparsely distributed and many of the inferred gains are distantly related on the trees (see Figure 1 and Figure S1). This pattern suggest that in most cases adding more salt tolerant taxa is likely to either increase the total number of inferred salt tolerance gains (if adding species that are not closely related to known halophytes) or maintain the number of inferred gains (if adding to a clade of known halophytes). However, there are some clades, for example in Cyperaceae and Amaranthaceae, with denser groups of halophytes, where the number of inferred gains relative to the number of halophytes is more likely to be reduced by adding more halophytes (Figure 1a,f). Similarly, removing identified halophytes with only low or seasonal tolerance to salinity could in some cases increase phylogenetic clustering of halophytes, reducing the number of inferred gains (if removing species that are not closely related to known halophytes), or break up some clades, possibly increasing the estimated number of gains. Based on the extant pattern of halophytes, our analysis suggests that salt tolerance has been gained at least 600 times in the 22 families analyzed. And if we assume that each of the other angiosperm families that contain halophytes also represents at least one independent gain of salt tolerance, there are likely to be 124 additional gains or more across the angiosperms.

Our results are consistent with the findings of two previous group-specific studies on phylogenetic patterns of salt tolerance [7, 8]. We infer over 100 gains of salt tolerance within the grass family subtree, and confirm that halophytes are more phylogenetically dispersed than expected under Brownian motion. We also demonstrate that the Amaranthaceae has relatively high numbers of species per inferred gain, indicating a more conserved pattern of salt tolerance evolution compared to the other families in the analysis. While we estimate that salt tolerance is significantly less clustered than Brownian motion in Amaranthaceae, this result appears to be driven by about one-third of the family with notably fewer halophytes than the rest of the tree (Figure 1a).

Given that salt tolerance may involve many anatomical, physiological and life history modifications, it may seem surprising that it has evolved so many times in such a wide range of lineages. However, it has been suggested that other stress tolerance strategies involving complex sets of ecophysiological traits have also evolved multiple times [17-20]. One explanation for how salt tolerance has evolved multiple times is that the

required physiological or anatomical changes can build on precursor traits acquired earlier in the history of the lineages. A well-studied example of how complex physiological traits can build on precursor traits is C_4 photosynthesis in the grasses [21]. In a few angiosperm families researchers have inferred many independent evolutionary origins of C_4 photosynthesis, a specialized form of photosynthesis often associated with arid-adapted lineages [19, 22, 23], which requires many biochemical and anatomical modifications. Lineages with a higher proportion of vascular bundle sheath cells have a higher frequency of evolution of C_4 photosynthesis, suggesting that some types of foliar anatomy facilitate the transition to C_4 [23]. Similarly, if salt tolerance builds on existing physiological or anatomical traits, then a lineage with these traits may have a higher likelihood of giving rise to halophytic species. For example, C_4 grass lineages are more likely to contain halophytes than C_3 lineages, possibly because C_4 photosynthesis allows more efficient water use and therefore limits the impact of salinity by reducing the uptake of ions and limiting the effects of osmotic stress [24].

Although the idea of evolutionary precursors may explain why some lineages develop salinity tolerance more often than others, the question remains: why are salt tolerant lineages often found as singletons on the phylogeny or in small clades? There are several broad explanations for this pattern, which are not mutually exclusive. One explanation is that the observed distribution of halophytes could reflect patterns of change in land salinity over time. Although some saline areas are long lived (e.g., coastal habitats), in some areas salinity can vary over small spatial scales or shift on a seasonal basis [25]. If lineages are rapidly responding to changing salinity, this could partly explain why we infer mostly shallow gains of salt tolerance that give rise to only one or a few extant halophytes. However, recent origins of saline habitats is unlikely to provide a general explanation for the multiple recent gains of salt tolerance in many families, because some saline habitats are stable over long evolutionary time periods, so should provide persistent habitat for saline specialists.

Another explanation for why there are so many small clades of halophytes is that salt tolerance may be a costly ecological strategy that is relatively easy to gain but difficult to maintain. For example, high plasticity could enable some lineages to transition into harsh or novel habitats over evolutionarily short time scales [26, 27]. However, maintaining a strategy like salt tolerance could be physiologically costly, for example due to the cost of producing osmoprotectants or increasing investment in reactive

oxygen species scavenging and antioxidant production (reviewed in [28]). The high physiological cost of salt tolerance could lead to increased extinction rates in halophytes, or high reversal rates if lineages that invest less in salt tolerance mechanisms have a competitive advantage. This scenario could lead to an extant pattern of many shallow gains of salt tolerance dispersed across the phylogeny. Some research suggests that the more salt tolerant a species, the less competitive it is in less saline or non-saline environment [29, 30], although the generality of these claims are disputed [31]. Reduced competitive ability may threaten the persistence of halophytes if land salinity subsides, and halophytes may not be ecologically competitive when transitioning back into a non-saline environment [32], which could lead to local extinction or the loss of salt tolerance. However, the lower competitive ability may not always be a direct result of salt tolerance [31], and high salinity tolerance may even confer a competitive advantage for some species in non-saline habitats. For example, salt cedars (*Tamarix*, Tamaricaceae) are highly salt tolerant, yet they are invasive in some non-saline and low-saline riparian habitats. Salt cedar populations are capable of displacing natives by using more water and excreting salt into the soil, creating a toxic environment for non-salt tolerant native species [33].

It is unlikely that either changes in land salinity patterns or the cost of salt tolerance can fully explain why salt tolerance has evolved many times and why halophytes are often found as singletons and in small clades. And it has been suggested that, in general, the transition into different habitats and the evolution of ecological traits may be highly context dependent [21]. Identifying the phylogenetic patterns of salt tolerance represents an important step towards understanding salt tolerance evolution. We hope that reporting results in the context of angiosperm families will be useful for more detailed studies in future on the environmental and physiological aspects of salt tolerance evolution in different lineages. In future it would be interesting to explore the role that related traits, order of trait acquisition, and climatic history have played in the observed patterns of salt tolerance evolution, as has been examined for C_4 photosynthesis [23] and freezing tolerance [34].

Conclusions

Salt tolerance in plants is an interesting case study in macroevolution [35]. Salt tolerance is an ecological strategy that often involves complex physiological features. Halophytes are rare, yet they are found in a diverse set of taxonomic groups. Our

analysis shows that in a range of angiosperm families, salt tolerance has been gained a surprising number of times and that these transitions are shallow and spread out near the tips of the phylogeny. This suggests that while the evolutionary pattern of salt tolerance varies across angiosperm families, it seems that salt tolerance can evolve frequently in many different genetic backgrounds. This result is intriguing, given how difficult it has to been to manipulate the salt tolerance of commercial crop varieties [31, 36], but the frequent evolution of salt tolerance may give hope that many plant lineages can build on existing physiological and anatomical traits to develop increased tolerance of environmental salinity.

Materials and Methods

Halophyte database

Our aim was to broadly investigate the patterns of salt tolerance distribution as an ecological strategy across angiosperms. We first compiled a list of known halophytes. Instead of differentiating halophytes based on specific traits or environmental conditions, we analyzed salt tolerance as a binary trait, categorizing plants as reported to tolerate salt (labeled as 1) or not reported as salt tolerant (labeled as 0). Analyzing salt tolerance as a binary character is the only practical approach for a broad scale comparative study since there are relatively few species for which we have information on specific levels of salt tolerance, and this approach also allowed us to study a wide variety of salt tolerant species. We started with a published list of approximately 2600 taxa observed in saline habitats [37]. We then searched the literature and added taxa from five additional halophyte lists that were published more recently [38-42] (See Additional files for details). These published lists included halophytes identified from field surveys and observational data. It is possible that some taxa included in these lists have low salinity tolerance, are only tolerant to limited exposure to salinity (e.g., seasonal salinity), or have experienced acclimation to salinity [43, 44]. For this study, we consider that these species have an underlying propensity for developing salt tolerance, and so their inclusion is useful in a broad study on the evolution of salt tolerance. The resulting list contained 4515 taxa reported to be salt tolerant (including infraspecific taxa). We then searched for synonyms and accepted names of each taxon in this list according to The Plant List [45] using the R package ‘taxonstand’ [46]. Because the taxonomic and phylogenetic analyses had different aims, we created separate lists for each analysis, which are described below.

Halophyte Diversity

Our first aim was to investigate the taxonomic distribution of halophytes across angiosperm families. Although families may represent lineages of different ages or evolutionary patterns, here they are used simply as a convenient taxonomic division of angiosperm diversity into defined groups. We identified 411 unique angiosperm families by checking the 413 families recognized by the Linear Angiosperm Phylogeny Group (LAPG III) [47] against the APG III website [48, 49] (two were found to have equivocal names, see Additional files). We then collected mean estimates of species numbers for each of the 411 families from the APG III website [49], totaling 276,000 species. Since these family size estimates are reported at the species level, we needed to compile a list of halophytic species. We selected only the unique set of accepted halophyte species names according to The Plant List [45]. We collapsed the names of accepted infraspecific taxa to the species level, counting a species as salt tolerant if one of its varieties or subspecies was listed as a halophyte. The resulting list contained 2852 unique halophyte species (see archived data for halophyte list). We then counted the number of known halophyte species in each angiosperm family.

Then, we tested if halophytes were distributed randomly across families. We tested whether the number of halophytes in each family followed a binomial distribution parameterized by family size and the probability of being salt tolerant equal to the observed proportion of halophytes over all the angiosperm families. We applied a G-test of independence to estimate the overall fit of the model to the angiosperm families, using the likelihood.test function in R package “Deducer” [50].

We then calculated the probability of observing the number of known halophytes for each family based on the binomial distribution. This probability allowed us to identify families with significantly more or fewer halophytes than expected by chance. If the probability of a family having the same number or more halophytes than observed is lower than 0.05, the family is considered to have significantly more halophytes than expected by chance. If the probability of a family having the same number or fewer halophytes than observed is lower than 0.05, the family is considered to have significantly fewer halophytes than expected by chance.

Evolutionary patterns

Phylogenies

Our second aim was to investigate the phylogenetic patterns of halophyte distribution. Because halophytes are rare, we did not analyze the distribution of salt tolerance across the entire angiosperm phylogeny, but focused on a subset of families that each contained many halophytes. Our main aims in family selection were to: (1) collect the largest families containing a sufficient number of halophytes to provide sufficient power for the analysis of evolutionary patterns; and (2) select families that were found to have more, fewer or within the expected range of halophytes in the taxonomic analysis, so as not to bias the analysis to families with a higher representation of salt tolerant taxa. All our analyses were conducted using phylogenetic information from the largest available phylogeny for angiosperms [12]. This phylogeny includes all appropriate angiosperm sequences on GenBank, including infraspecific taxa, which covers approximately 20% of angiosperm species.

We first assigned each tip in the phylogeny to a family according to GenBank taxonomy using the TaxoGB function in the R package "BoSSA" [51], and recorded the total number of tips (terminal taxa in the phylogeny) associated with each family. We then determined which tips in the phylogeny were halophytes, based on whether the tip name was included in either the halophyte names presented in the original publications or the accepted names found in the synonymy search (total 5030 halophyte names, see archived data files). We chose this identification method because the published phylogeny includes all angiosperm sequences on GenBank and is not restricted to the taxonomic classification on The Plant List [45]. We did not collapse infraspecific taxa to the species level for this analysis since the published phylogeny includes infraspecific taxa. We also identified whether the tips associated with each family were monophyletic in the phylogeny. In order to restrict the analysis to families with sufficiently large phylogenies and more than a few halophytes, we considered only families with 25 or more taxa included in the Smith *et al*. [12] phylogeny, of which at least six were recognized halophytes.

Under these criteria, we selected 22 families, including families with more halophytes than expected by chance, families with fewer halophytes than expected, and also families that fell within the expected number of halophytes for the family size. We first extracted sixteen families that were monophyletic in the Smith *et al*. [12] phylogeny.

For each monophyletic family we extracted the family subtree from the phylogeny, which included all tips in the Smith *et al*. [12] phylogeny belonging to that family according to GenBank taxonomy. Next we extracted subtrees for families that met our selection criteria but were not strictly monophyletic in the Smith *et al.* [12] tree. For these families, we extracted a monophyletic family subtree by removing a small number of taxa that were assigned to the target family in GenBank taxonomy but did not fall into that family clade in the published phylogeny. In some cases we also excluded a small number of taxa within the target family clade that were assigned to a different family (see Additional files for details on the names of taxa excluded from each family subtree). We then removed tips from the family subtrees that were not identified with standardized genus and species epithets. We excluded any tips with labels that included the taxonomic epithets “af”, “aff”, “cf” or “sp”. We also removed tips that represented hybrid taxa by identifying tip labels that included one genus and two specific epithets, as well as the word “hybrid” or where the two species names were separated by “x”. All family subtrees were extracted and analyzed with equal branch lengths. Polytomies in the family subtrees were randomly resolved using the multi2di function in the R package ‘ape’ [52].

Metrics for analyzing phylogenetic patterns

For each family subtree we used two metrics to assess the phylogenetic pattern of halophytes within the family. Specifically, we aimed to test whether any of these families showed the same evolutionary pattern of salt tolerance as the grass family, having (1) many shallow inferred origins of salt tolerance, near the tips of the phylogeny, which gave rise to small clades of halophytes, and (2) origins that were spread across the phylogeny, occurring in many different lineages [8]. To detect these patterns in our sample of families, we used two metrics: the number of tips per origin (NoTO) and the sum of sister clade differences, a measure of phylogenetic clustering (SSCD) [53, 54].

The number of tips per origin (NoTO) metric is used to test whether, given the number of halophyte taxa in the tree (tips), there are significantly more inferred origins of salt tolerance, and thus smaller clades of halophytes, than we would expect under a Brownian motion model of trait evolution. It is possible that salt tolerance has been gained and lost multiple times and that salt tolerant lineages have since gone extinct [35]. Here we only infer gains of salt tolerance that lead to extant salt tolerant species,

as our aim is to infer the minimum number of independent gains needed to explain the extant phylogenetic distribution of halophytes. Our aim was to compare the observed taxonomic and phylogenetic distribution of halophytes to a null model of trait evolution using trait reconstruction techniques. To estimate the NoTO for each family subtree, we inferred the minimum number of gains and losses of salt tolerance required to explain the observed topological distribution of halophyte tips, and then calculated the average number of halophyte tips per inferred gain. A shallow, scattered distribution of halophytes, where most halophyte species arise from gains near the tips of the phylogeny and most gains lead to only a few halophyte species, will have a low value for NoTO.

To generate a null distribution of the expected number of gains given a number of known halophytes, we used established methods to simulate salt tolerance as a continuous trait on each family subtree using a Brownian motion model [54, 55]. We then used an appropriate threshold to convert the continuous trait to a binary one, such that the number of halophyte tips in the simulated tree was equal to the number of identified halophytes in the family subtree. We repeated this process 1000 times to generate a null distribution of NoTO values, specific to the observed number of halophytes and the size of the subtree for comparison with the observed NoTO value for each family subtree. To generate a *p*-value for each family, we calculated the proportion of simulated trees that had a NoTO value lower than or equal to the observed. *P*-values less than or equal to 0.05 represent significantly smaller clades of halophytes than expected under Brownian motion.

The sum of sister clade differences (SSCD) metric describes the degree of phylogenetic clustering of halophytes. We used the method of calculating SSCD described by Fritz and Purvis [54]. Each tip was coded as 1 if it was on the halophyte list and 0 if it was not. Each internal node in the family subtree was assigned a trait state using the mean of the descendant node or tip states (e.g., if one descendant was state 1 and the other was 0, the node value was 0.5). The SSCD was calculated as the sum of the absolute difference between trait states of each pair of sister nodes or tips over the whole tree. If gains of salt tolerance were scattered across the phylogeny, each giving rise to one or few halophyte species in a small clade of salt tolerant taxa, we would expect a large SSCD value. We compared the observed SSCD of each family subtree to the SSCD for 1000 traits generated under Brownian motion on the subtree, using the same method for

generating the null distribution for the NoTO value. The *p*-value was the proportion of simulated trees that have higher SSCD values than the observed. *P*-values less than or equal to 0.05 indicated that salt tolerance is significantly scattered on the phylogeny compared to a Brownian motion trait.

We conducted correlation tests to assess whether estimates of NoTO and SSCD were influenced by incomplete sampling in the family subtrees or by the proportion of halophytic taxa in the family subtrees. Since NoTO, SSCD, and sampling values represent proportions, we used a non-parametric correlation test, Kendall's tau [56].

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

C.M. collected the data and performed the analyses. X.H. developed the methods. L.B. proposed the study, and L.B. and C.M. designed the study. All authors wrote and reviewed the manuscript.

Availability of Supporting Data

The data set supporting the results of this article is available in the Dryad repository. http://dx.doi.org/10.5061/dryad.n64kk.

Acknowledgments

We would like to thank Thomas H. Bennett for assistance with data collection and project design, Robert Lanfear for methodological assistance and Timothy J. Flowers for his encouragement and helpful feedback.

References

1. Glenn EP, Brown JJ, Blumwald E: **Salt tolerance and crop potential of halophytes**. *Crit Rev Plant Sci* 1999, **18**:227–255.

2. Flowers TJ, Colmer TD: **Salinity tolerance in halophytes**. *New Phytol* 2008, **179**:945–963.

3. Greenway H, Munns R: **Mechanisms of salt tolerance in nonhalophytes**. *Annu Rev Plant Phys* 1980, **31**:149–190.

4. Flowers TJ: **Improving crop salt tolerance**. *J Exp Bot* 2004, **55**:307.

5. Flowers T, Troke PF, Yeo AR: **The mechanism of salt tolerance in halophytes**. *Annu Rev Plant Phys* 1977, **28**:89–121.

6. Flowers TJ, Galal HK, Bromham L: **Evolution of halophytes: multiple origins of salt tolerance in land plants**. *Functional Plant Biol* 2010, **37**:604.

7. Kadereit G, Ackerly D, Pirie MD: **A broader model for C_4 photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae s.s.).** *Proc R Soc Lond B Biol Sci* 2012, **279**:3304–3311.

8. Bennett TH, Flowers TJ, Bromham L: **Repeated evolution of salt-tolerance in grasses**. *Biol lett* 2013, **9**:20130029–20130029.

9. Flowers TJ: **Physiology of halophytes**. *Plant Soil* 1985, **89**:41–56.

10. Tester M, Davenport R: **Na^+ tolerance and Na^+ transport in higher plants**. *Ann Bot* 2003, **91**:503–527.

11. Shi S, Huang Y, Zeng K, Tan F, He H, Huang J, Fu Y: **Molecular phylogenetic analysis of mangroves: independent evolutionary origins of vivipary and salt secretion**. *Mol Phylogenet Evol* 2005, **34**:159–166.

12. Smith SA, Beaulieu JM, Stamatakis A, Donoghue MJ: **Understanding angiosperm diversification using small and large phylogenetic trees**. *Am J Bot* 2011, **98**:404–414.

13. Waisel Y: *The biology of halophytes*. London: Academic Press; 1972.

14. Joseph S, Bhave M, Miller JT, Murphy DJ: **Rapid identification of *Acacia* species with potential salt tolerance by using nuclear ribosomal DNA markers**. *Sustainable Agriculture Research* 2013, **2**.

15. Saslis-Lagoudakis CH, Hua X, Bui E, Moray C, Bromham L: **Predicting species' tolerance to salinity and alkalinity using distribution data and geochemical modelling: a case study using Australian grasses.** *Ann Bot* 2015, **115**:343–351.

16. Bui EN, Thornhill A, Miller JT: **Salt- and alkaline-tolerance are linked in *Acacia***. *Biol lett* 2014, **10**:20140278–20140278.

17. Crayn DM, Winter K, Smith JAC: **Multiple origins of crassulacean acid metabolism and the epiphytic habit in the Neotropical family Bromeliaceae.** *Proc Natl Acad Sci USA* 2004, **101**:3703–3708.

18. Anacker BL, Whittall JB, Goldberg EE, Harrison SP: **Origins and consequences of serpentine endemism in the California flora**. *Evolution* 2010, **65**:365–376.

19. Sage RF, Christin PA, Edwards EJ: **The C_4 plant lineages of planet Earth**. *J Exp Bot* 2011, **62**:3155–3169.

20. Cappa JJ, Pilon-Smits EAH: **Evolutionary aspects of elemental hyperaccumulation**. *Planta* 2014, **239**:267–275.

21. Edwards EJ, Donoghue MJ: **Is it easy to move and easy to evolve? Evolutionary accessibility and adaptation**. *J Exp Bot* 2013, **64**:4047–4052.

22. Edwards EJ, Smith SA: **Phylogenetic analyses reveal the shady history of C_4 grasses**. *Proc Natl Acad Sci USA* 2010, **107**:2532–2537.

23. Christin P-A, Osborne CP, Chatelet DS, Columbus JT, Besnard G, Hodkinson TR, Garrison LM, Vorontsova MS, Edwards EJ: **Anatomical enablers and the evolution of C_4 photosynthesis in grasses.** *Proc Natl Acad Sci* 2013, **110**:1381–1386.

24. Bromham L, Bennett TH: **Salt tolerance evolves more frequently in C_4 grass lineages**. *J Evol Biol* 2014, **27**:653–659.

25. Bui EN: **Soil salinity: A neglected factor in plant ecology and biogeography**. *J Arid Environ* 2013, **92**(C):14–25.

26. Antonovics J, Bradshaw AD, Turner RG: **Heavy metal tolerance in plants**. *Adv Ecol Res* 1971, **7**:1–85.

27. Richards CL, Walls RL, Bailey JP, Parameswaran R, George T, Pigliucci M: **Plasticity in salt tolerance traits allows for invasion of novel habitat by Japanese knotweed s. l. (*Fallopia japonica* and *F.xbohemica*, Polygonaceae)**. *Am J Bot* 2008, **95**:931–942.

28. Bose J, Rodrigo-Moreno A, Shabala S: **ROS homeostasis in halophytes in the context of salinity stress tolerance**. *J Exp Bot* 2014, **65**:1241–1257.

29. Egan TP, Ungar IA: **Competition between *Salicornia europaea* and *Atriplex prostrata* (Chenopodiaceae) along an experimental salinity gradient**. *Wetlands Ecology and Management* 2001, **9**:457–461.

30. Greenwood ME, MacFarlane GR: **Effects of salinity on competitive interactions between two Juncus species**. *Aquat Bot* 2009, **90**:23–29.

31. Rozema J, Schat H: **Salt tolerance of halophytes, research questions reviewed in the perspective of saline agriculture**. *Environ Exp Bot* 2013, **92**:83–95.

32. Ungar IA: **Are biotic factors significant in influencing the distribution of halophytes in saline habitats?** *Bot Rev* 1998, **64**:176–199.

33. Ladenburger CG, Hild AL, Kazmer DJ, Munn LC: **Soil salinity patterns in *Tamarix* invasions in the Bighorn Basin, Wyoming, USA**. *J Arid Environ* 2006, **65**:111–128.

34. Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG,

McGlinn DJ, O'Meara BC, Moles AT, Reich PB, Royer DL, Soltis DE, Stevens PF, Westoby M, Wright IJ, Aarssen L, Bertin RI, Calaminus A, Govaerts R, Hemmings F, Leishman MR, Oleksyn J, Soltis PS, Swenson NG, Warman L, Beaulieu JM: **Three keys to the radiation of angiosperms into freezing environments**. *Nature* 2014, **506**:89–92.

35. Bromham L: **Macroevolutionary patterns of salt tolerance in angiosperms.** *Ann Bot* 2015, **115**:333–341.

36. Rozema J, Flowers T: **Crops for a salinized world.** *Science* 2008, **322**:1478–1480.

37. Menzel U, Lieth H: **Halophyte database version 2.0**. In *Cash crop halophytes: recent studies*. Dordrecht: Kluwer Academic Publishers; 2003:221–250.

38. Khan MA, Qaiser M: **Halophytes of Pakistan: characteristics, distribution and potential economic usages**. *Sabkha ecosystems* 2006:129–153.

39. Guvensen A, Gork G, Özturk M: **An overview of the halophytes in Turkey**. *Sabkha ecosystems* 2006:9–30.

40. Dagar J, Gurbachan S: **Biodiversity of saline and waterlogged environments : documentation, utilization and management**. *NBA Scientific Bulletin Number - 9,* National Biodiversity Authority, Chennai, TamilNadu, India*;* 2007:78.

41. Özturk M, Guvensen A, Sakçali S, Gork G: **Halophyte plant diversity in the Irano-Turanian phytogeographical region of Turkey**. In *Biosaline Agriculture and High Salinity Tolerance*. Basel: Birkhäuser; 2008:141–155.

42. Zhao K, Song J, Feng G, Zhao M, Liu J: **Species, types, distribution, and economic potential of halophytes in China**. *Plant Soil* 2011, **342**:495–509.

43. Djanaguiraman M, Sheeba JA, Shanker AK, Devi DD, Bangarusamy U: **Rice can acclimate to lethal level of salinity by pretreatment with sublethal level of salinity through osmotic adjustment**. *Plant Soil* 2006, **284**:363–373.

44. Pandolfi C, Mancuso S, Shabala S: **Physiology of acclimation to salinity stress in pea (*Pisum sativum*)**. *Environ Exp Bot* 2012, **84**:44–51.

45. **The Plant List**. 2010. Version 2 [http://www.theplantlist.org/]

46. Cayuela L, Granzow-de la Cerda Í, Albuquerque FS, Golicher DJ: **Taxonstand: An R package for species names standardisation in vegetation databases**. *Methods Ecol Evol* 2012, **3**:1078–1083.

47. Haston E, Richardson JE, Stevens PF, Chase MW, Harris DJ: **The Linear Angiosperm Phylogeny Group (LAPG) III: a linear sequence of the families in APG III**. *Bot J Linn Soc* 2009, **161**:128–131.

48. APG III: **An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III.** *Bot J Linn Soc* 2009, **161**:105–121.

49. Stevens, PF. **Angiosperm Phylogeny Website**. 2001. Version 12, July 2012. [http://www.mobot.org/MOBOT/research/APweb/]

50. Fellows I: **Deducer: A data analysis GUI for R**. *J Stat Softw* 2012, **49**:1–15.

51. Lefeuvre P: **BoSSA: a Bunch of Structure and Sequence Analysis**.

52. Paradis E, Claude J, Strimmer K: **APE: analyses of phylogenetics and evolution in R language**. *Bioinformatics* 2004, **20**:289.

53. Felsenstein J: **Phylogenies and the comparative method**. *Am Nat* 1985:1–15.

54. Fritz SA, Purvis A: **Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits.** *Conserv Biol* 2010, **24**:1042–1051.

55. Felsenstein J: **Using the quantitative genetic threshold model for inferences between and within species**. *Philos Trans R Soc Lond B Biol Sci* 2005, **360**:1427–1434.

56. Kendall MG: **A new measure of rank correlation**. *Biometrika* 1938, **30**:81–93.

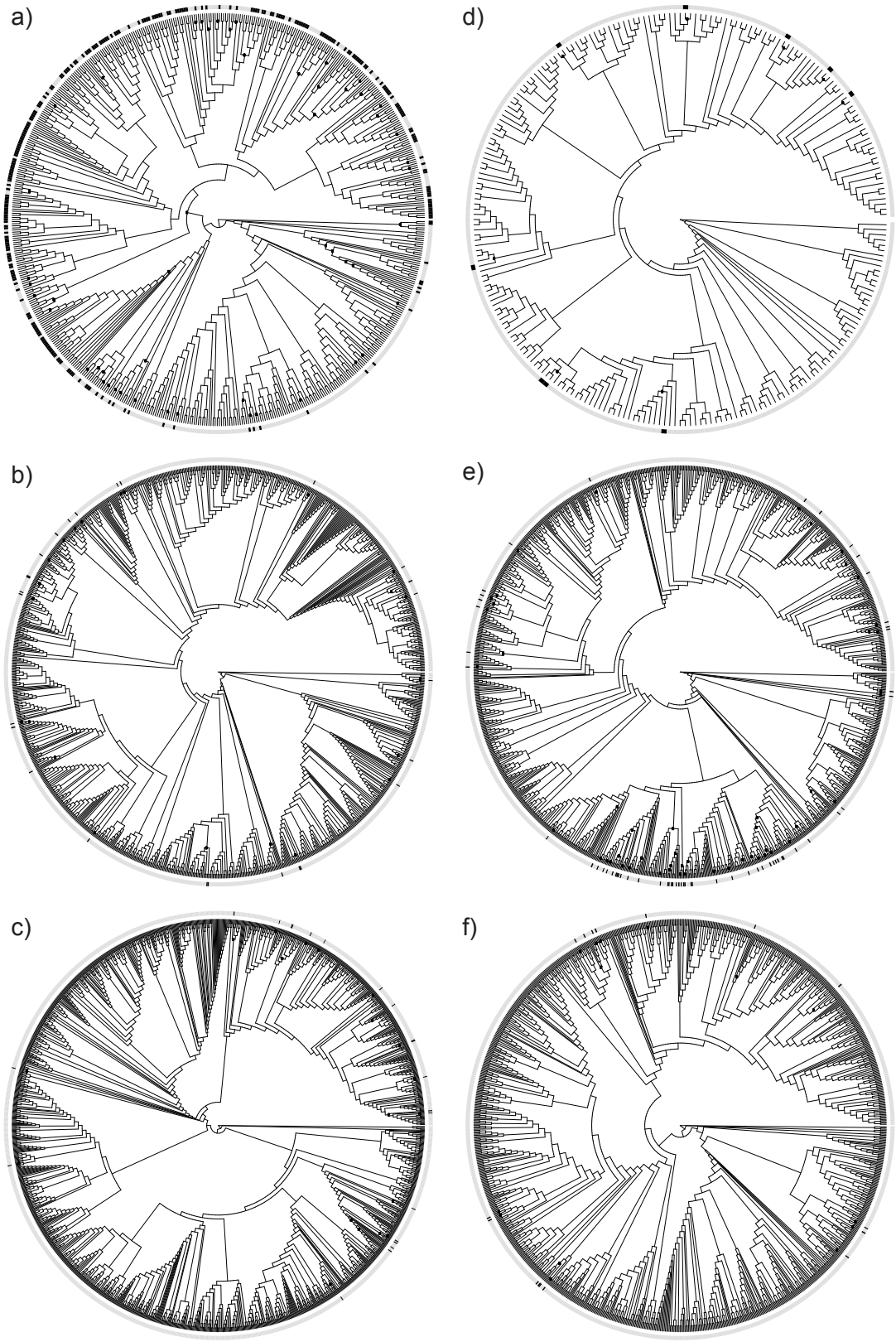
57. FitzJohn RG: **Diversitree: comparative phylogenetic analyses of diversification in R**. *Methods Ecol Evol* 2012, **3**:1084-1092.

58. Moray C, Hua X, Bromham L: **Data from: Salt tolerance is evolutionarily labile in a diverse set of angiosperm families**. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.n64kk.

Table 1: Results of taxonomic and phylogenetic analyses for a sample of 22 angiosperm families. Family and order names are based on APG III [48]. Family size is the mean estimated number of species in the family reported on the APG III website [49]. The halophytes column lists the number of known halophytes species in each family. Family subtree size represents the number of taxa in the phylogenetic tree used for analysis, and halophytes in subtree is the number of known halophytes included in each family subtree. The halophytes sampled in subtree represents the percent of known halophytes that are present in each family subtree. The taxonomic pattern column identifies families with more or fewer halophytes than expected by chance based on the taxonomic analysis (see Methods). The results of the metrics used to distinguish evolutionary patterns of salt tolerance are presented. For the number of tips per origin (NoTO), *p*-values represent whether the average number of halophytes arising from each inferred gain of salt tolerance is smaller expected under Brownian motion ($p < 0.05$). For the sum of sister clade differences (SSCD), *p*-values represent whether halophytes are less clustered than expected under Brownian motion ($p < 0.05$). Test statistics that are significantly different to the null model are presented in bold. Significant results for Tamaricaceae are italicized to highlight that this is the only significantly conserved pattern of salt tolerance, where significantly more tips per gain and a significantly smaller SSCD.

Order	Family	Family size	Known Halophytes	Halophytes in family (%)	Taxonomic pattern	Family subtree size	Species in subtree (%)	Halophytes in subtree	Halophytes in subtree (%)	Halophytes sampled in subtree (%)	Inferred origins	NoTO	NoTO (*p*)	SSCD (*p*)
Apiales	Apiaceae	3780	33	0.9		1082	28.6	26	2.4	78.8	22	1.2	**0.00**	**0.00**
Arecales	Arecaceae	2361	35	1.5	more	415	17.6	19	4.6	54.3	15	1.3	0.06	**0.00**
Asterales	Asteraceae	23600	267	1.1		4618	19.6	97	2.1	36.3	87	1.1	**0.00**	**0.00**
-	Goodeniaceae	430	6	1.4		69	16.0	6	8.7	100	6	1.0	0.09	**0.01**
Brassicales	Brassicaceae	3710	38	1.0		1355	36.5	21	1.5	55.3	19	1.1	**0.00**	**0.00**
Caryophyllales	Amaranthaceae	2275	507	22.3	more	613	26.9	262	42.7	51.7	54	4.9	0.16	**0.00**
-	Tamaricaceae	90	55	61.1	more	42	46.7	29	69.0	52.7	1	29.0	***1.00***	***1.00***
Cucurbitales	Cucurbitaceae	960	14	1.5		247	25.7	9	3.6	64.3	8	1.1	0.14	**0.02**
Ericales	Primulaceae	2590	14	0.5	fewer	546	21.1	8	1.5	57.1	5	1.6	0.65	0.55
Fagales	Casuarinaceae	95	12	12.6	more	88	92.6	12	13.6	100	7	1.7	0.46	0.08
Gentianales	Rubiaceae	13150	13	0.1	fewer	1393	10.6	7	0.5	53.8	7	1.0	0.09	**0.01**
Lamiales	Acanthaceae	4000	18	0.5	fewer	498	12.5	9	1.8	50.0	5	1.8	0.75	0.54
-	Lamiaceae	7173	27	0.4	fewer	941	13.1	14	1.5	51.9	11	1.3	0.14	0.05
Malpighiales	Euphorbiaceae	5735	42	0.7	fewer	1047	18.3	16	1.5	38.1	14	1.1	**0.03**	**0.00**
-	Rhizophoraceae	149	19	12.8	more	40	26.8	18	45.0	94.7	6	3.0	0.52	0.71
Myrtales	Combretaceae	500	12	2.4	more	25	5.0	8	32.0	66.7	6	1.3	0.23	0.25
-	Lythraceae	620	21	3.4	more	119	19.2	14	11.8	66.7	8	1.8	0.53	0.46
-	Myrtaceae	4620	47	1.0		612	13.2	20	3.3	42.6	19	1.1	**0.00**	**0.00**
Poales	Cyperaceae	5430	121	2.2	more	1087	20.0	57	5.2	47.1	52	1.1	**0.00**	**0.00**
-	Juncaceae	430	22	5.1	more	124	28.8	12	9.7	54.5	8	1.5	0.31	0.45
-	Poaceae	11160	335	3.0	more	2291	20.5	173	7.6	51.6	127	1.4	**0.00**	**0.00**
Rosales	Rosaceae	2520	9	0.4	fewer	1010	40.1	8	0.8	88.9	8	1.0	**0.05**	**0.05**

Figure 1: Family subtrees for a sample of six of the families analyzed with significant NoTO and/or SSCD values. Inferred gains of salt tolerance (see Methods) are marked on each family with black circles. Tips in the subtrees identified as halophytes are marked in black in the ring around the subtree. The subtrees represent a) Amaranthaceae, b) Apiaceae, c) Brassicaceae, d) Cucurbitaceae, e) Cyperaceae, and f) Euphorbiaceae. All 22 subtrees included in the analysis are presented in Figure S1. Subtree plots were created using the “Diversitree” package (57).


a)
d)
b)
e)
c)
f)

Supplemental material for "**Salt tolerance is evolutionarily labile in a diverse set of angiosperm families**"

C. Moray, X. Hua, L. Bromham

1. Composition of halophyte list

We started with a list of halophytes from Menzel and Lieth (2003), then added halophytes included in lists of salt tolerant species from more recent publications. We then verified the list for synonymy, as described in the Materials and Methods section. Here we outline which species we included in our halophyte list from each source, since the definitions and terminology used to identify salt tolerant taxa differs between sources. Because information on specific levels of salinity tolerance is rare, in general we included species that were listed as salt tolerant based on observational evidence (that is, reported as being able to complete their life cycle under saline conditions). This means that we may have included species with relatively low levels of salt tolerance, or tolerant to external and occasional exposure to salinity (i.e., salt spray), so it is possible that not all species in our compiled halophyte list are able to grow in highly salt-affected soil.

We added taxa identified as halophytes from studies in Turkey (Guvensen *et al.*, 2006; Özturk *et al.*, 2008), China (Zhao *et al.*, 2010), Pakistan (Khan & Qaiser, 2006). From Dagar and Gurbachan (2007), we added all species listed as true halophytes, facultative halophytes and glycophytes/transitional halophytes, since their definition of glycophytes/transitional halophytes includes species that are able to grow in saline soils.

2. Identification of Angiosperm families for taxonomic analysis

We identified 411 unique families based on the APG website (Stevens, 2001), whereas some sources have identified 413 (Haston *et al.*, 2009). For this analysis we started with

the 413 families listed by the LAPG III (Haston *et al.*, 2009) and checked all family names against the APG III website (Stevens, 2001). During this search we found that Aristolochiaceae and Lactoridaceae are considered one family by the APG III website and that Buxaceae and Haptanthaceae are also considered synonyms (Stevens, 2001). Here we considered these families as synonyms, reducing the number of angiosperm families considered in this analysis from 413 to 411. We also recognized Ripogonaceae (Haston *et al.*, 2009) as an alternative spelling of Rhipogonaceae (APG III, 2009).

3. Extraction of non-monophyletic family subtrees

Based on the selection criteria chosen for the phylogenetic analysis (see Methods), we needed to extract some subtrees for families that were not monophyletic in the published angiosperm phylogeny (Smith *et al.*, 2011). In general we extracted all tips associated with each target family, excluding a small number of tips that were either nested within clades of other families or tips from other families that were nested within the clade of the target family. Here we list details on each non-monophyletic family subtree, referring to specific tip numbers associated with the original published phylogeny. For the Asteraceae subtree, we excluded one Asteraceae tip that was nested within the Campanulaceae clade (tip number 10079). For Brassicaceae we excluded five tips that were in the Capparaceae clade (42135, 42136, 42145, 42168, 42171). For Euphorbiaceae we excluded four Peraceae tips that were nested in the Euphorbiaceae clade (36371:36374). For Lamiaceae we excluded one tip that was in the Verbenaceae clade (21536), and one Orobanchaceae (24129) that was in the Lamiaceae clade. For Rosaceae we excluded one tip that was in Ranunculaceae (27875). For Rubiaceae we excluded one tip that was in the Clusiaceae clade (35882).

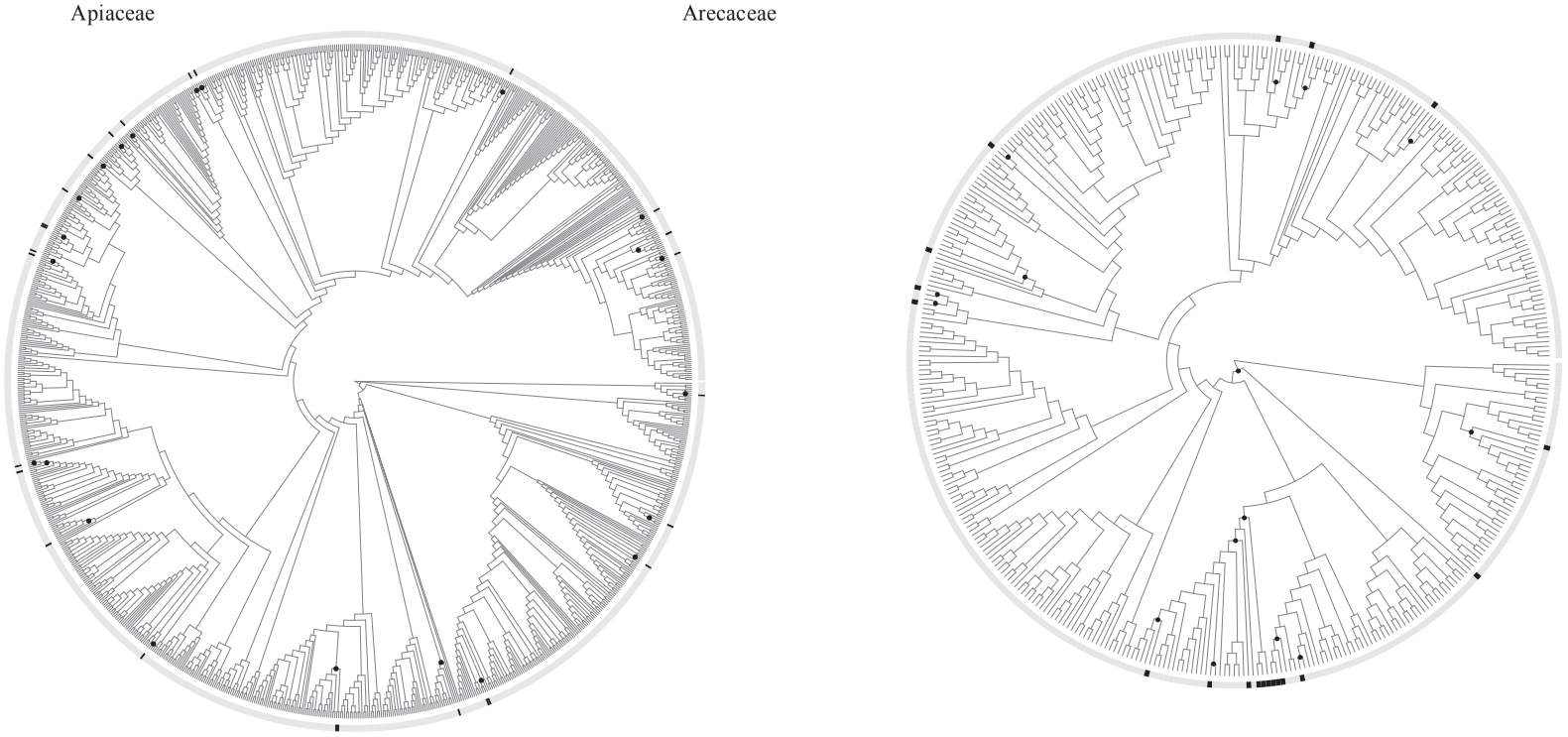
Table S1: Results of taxonomic analysis identifying families with more or fewer halophytes than expected for the 146 families with one or more halophytes. Family and order names come from the APG III website (Stevens, 2001). Genera and species represent the mean number of estimated genera and species in each family according to the APG III website (Stevens, 2001). Observed number and percentage of halophytes are based on the family affiliation of each accepted species in the halophyte list according to The Plant List (The Plant List, 2010). For Zosteraceae there were more observed halophytes (17) than estimated species in the family (14) since the mean species estimates come from the APG III website and the accepted species names in the halophyte list were confirmed with The Plant List (2010). For the analysis we considered Zosteraceae to have 100% halophytes, since the test is not valid when there are more halophytes than total species. *P*-values represent whether each family has more or fewer halophytes than expected under a binomial distribution (see Methods).

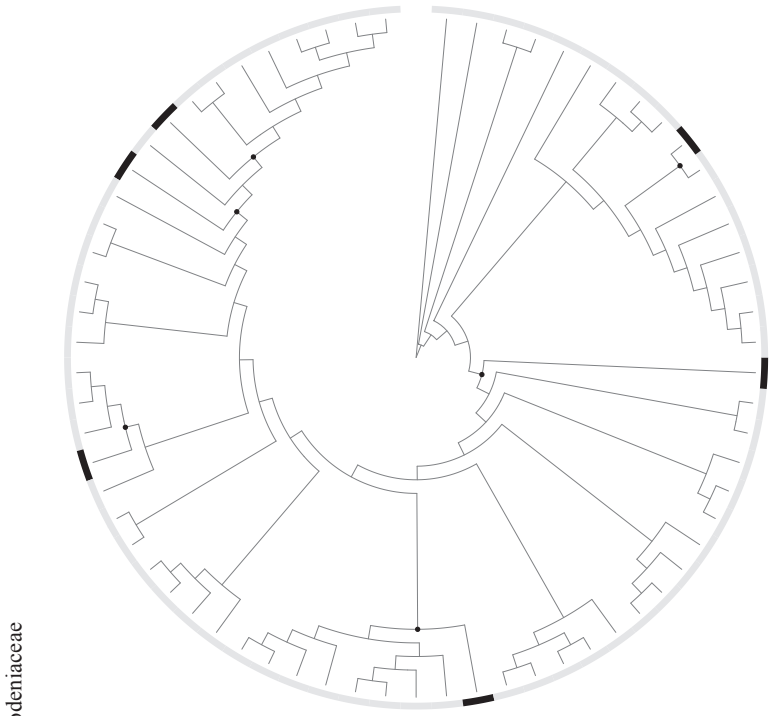
Order name (APG III)	Family name (APG III)	Species	Observed # halophytes	Observed % halophytes	*p*-value fewer	*p*-value more	Pattern
Alismatales	Alismataceae	88	3	3.4	0.99	0.06	
-	Aponogetonaceae	43	1	2.3	0.93	0.36	
-	Araceae	4759	4	0.1	0.00	1.00	fewer
-	Butomaceae	1	1	100.0	1.00	0.01	more
-	Cymodoceaceae	16	15	93.8	1.00	0.00	more
-	Hydrocharitaceae	116	22	19.0	1.00	0.00	more
-	Juncaginaceae	15	3	20.0	1.00	0.00	more
-	Posidoniaceae	9	3	33.3	1.00	0.00	more
-	Potamogetonaceae	102	6	5.9	1.00	0.00	more
-	Ruppiaceae	6	2	33.3	1.00	0.00	more
-	Zosteraceae	14	14	100.0	1.00	0.00	more
Apiales	Apiaceae	3780	33	0.9	0.19	0.85	
-	Araliaceae	1450	3	0.2	0.00	1.00	fewer
Arecales	Arecaceae	2361	35	1.5	0.98	0.02	more
Asparagales	Amaryllidaceae	1605	15	0.9	0.41	0.68	
-	Asparagaceae	2480	22	0.9	0.28	0.79	
-	Iridaceae	2025	10	0.5	0.01	1.00	fewer
-	Orchidaceae	22075	3	0.0	0.00	1.00	fewer
-	Xanthorrhoeaceae	900	3	0.3	0.02	1.00	fewer
Asterales	Asteraceae	23600	267	1.1	0.94	0.07	
-	Calyceraceae	60	1	1.7	0.87	0.46	
-	Goodeniaceae	430	6	1.4	0.84	0.29	
Brassicales	Bataceae	2	2	100.0	1.00	0.00	more
-	Brassicaceae	3710	38	1.0	0.53	0.54	
-	Capparaceae	480	10	2.1	0.99	0.03	more
-	Cleomaceae	300	7	2.3	0.99	0.04	more
-	Resedaceae	75	5	6.7	1.00	0.00	more
-	Salvadoraceae	11	4	36.4	1.00	0.00	more
Caryophyllales	Aizoaceae	2035	45	2.2	1.00	0.00	more
-	Amaranthaceae	2275	507	22.3	1.00	0.00	more
-	Anacampserotaceae	32	1	3.1	0.96	0.28	
-	Basellaceae	19	2	10.5	1.00	0.02	more
-	Cactaceae	1866	11	0.6	0.03	0.98	fewer
-	Caryophyllaceae	2200	25	1.1	0.73	0.34	
-	Didiereaceae	16	2	12.5	1.00	0.01	more
-	Frankeniaceae	90	15	16.7	1.00	0.00	more

Order name (APG III)	Family name (APG III)	Species	Observed # halophytes	Observed % halophytes	*p*-value fewer	*p*-value more	Pattern
-	Gisekiaceae	5	1	20.0	1.00	0.05	
-	Halophytaceae	1	1	100.0	1.00	0.01	more
-	Lophiocarpaceae	6	1	16.7	1.00	0.06	
-	Molluginaceae	87	4	4.6	1.00	0.01	more
-	Nyctaginaceae	395	9	2.3	0.99	0.02	more
-	Plumbaginaceae	836	62	7.4	1.00	0.00	more
-	Polygonaceae	1110	40	3.6	1.00	0.00	more
-	Portulacaceae	70	11	15.7	1.00	0.00	more
-	Sarcobataceae	2	1	50.0	1.00	0.02	more
-	Simmondsiaceae	1	1	100.0	1.00	0.01	more
-	Stegnospermataceae	3	1	33.3	1.00	0.03	more
-	Talinaceae	27	2	7.4	1.00	0.03	more
-	Tamaricaceae	90	55	61.1	1.00	0.00	more
Celastrales	Celastraceae	1400	8	0.6	0.05	0.98	fewer
Ceratophyllales	Ceratophyllaceae	6	1	16.7	1.00	0.06	
Commelinales	Commelinaceae	652	4	0.6	0.20	0.90	
-	Pontederiaceae	33	3	9.1	1.00	0.00	more
Cornales	Loasaceae	265	1	0.4	0.24	0.94	
Cucurbitales	Cucurbitaceae	960	14	1.5	0.92	0.13	
Dilleniales	Dilleniaceae	355	1	0.3	0.12	0.97	
Dipsacales	Caprifoliaceae	890	2	0.2	0.01	1.00	fewer
Ericales	Ebenaceae	548	4	0.7	0.33	0.82	
-	Ericaceae	3995	1	0.0	0.00	1.00	fewer
-	Lecythidaceae	310	4	1.3	0.78	0.40	
-	Primulaceae	2590	14	0.5	0.01	1.00	fewer
-	Sapotaceae	1100	2	0.2	0.00	1.00	fewer
-	Tetrameristaceae	5	1	20.0	1.00	0.05	
Fabales	Fabaceae	19500	243	1.2	1.00	0.00	more
-	Polygalaceae	965	3	0.3	0.01	1.00	fewer
-	Surianaceae	8	1	12.5	1.00	0.08	
Fagales	Betulaceae	145	1	0.7	0.56	0.78	
-	Casuarinaceae	95	12	12.6	1.00	0.00	more
Gentianales	Apocynaceae	4555	43	0.9	0.31	0.74	
-	Gentianaceae	1655	13	0.8	0.20	0.87	
-	Loganiaceae	420	1	0.2	0.07	0.99	
-	Rubiaceae	13150	13	0.1	0.00	1.00	fewer
Geraniales	Geraniaceae	805	1	0.1	0.00	1.00	fewer
Lamiales	Acanthaceae	4000	18	0.5	0.00	1.00	fewer
-	Bignoniaceae	800	9	1.1	0.69	0.44	
-	Lamiaceae	7173	27	0.4	0.00	1.00	fewer
-	Linderniaceae	195	2	1.0	0.67	0.60	
-	Orobanchaceae	2060	17	0.8	0.21	0.85	
-	Pedaliaceae	70	1	1.4	0.84	0.52	
-	Phrymaceae	188	4	2.1	0.95	0.13	
-	Plantaginaceae	1900	34	1.8	1.00	0.00	more
-	Scrophulariaceae	1800	15	0.8	0.24	0.83	
-	Verbenaceae	918	15	1.6	0.97	0.06	
Laurales	Hernandiaceae	55	2	3.6	0.98	0.11	
-	Lauraceae	2500	2	0.1	0.00	1.00	fewer
Liliales	Colchicaceae	245	2	0.8	0.54	0.72	
-	Liliaceae	610	1	0.2	0.01	1.00	fewer
Magnoliales	Annonaceae	2220	1	0.0	0.00	1.00	fewer
Malpighiales	Bonnetiaceae	35	1	2.9	0.95	0.30	
-	Chrysobalanaceae	460	1	0.2	0.05	0.99	fewer
-	Clusiaceae	595	2	0.3	0.06	0.98	
-	Elatinaceae	35	7	20.0	1.00	0.00	more
-	Euphorbiaceae	5735	42	0.7	0.01	0.99	fewer
-	Hypericaceae	560	1	0.2	0.02	1.00	fewer
-	Linaceae	300	4	1.3	0.80	0.37	
-	Phyllanthaceae	1745	9	0.5	0.02	0.99	fewer
-	Putranjivaceae	210	1	0.5	0.36	0.89	
-	Rhizophoraceae	149	19	12.8	1.00	0.00	more

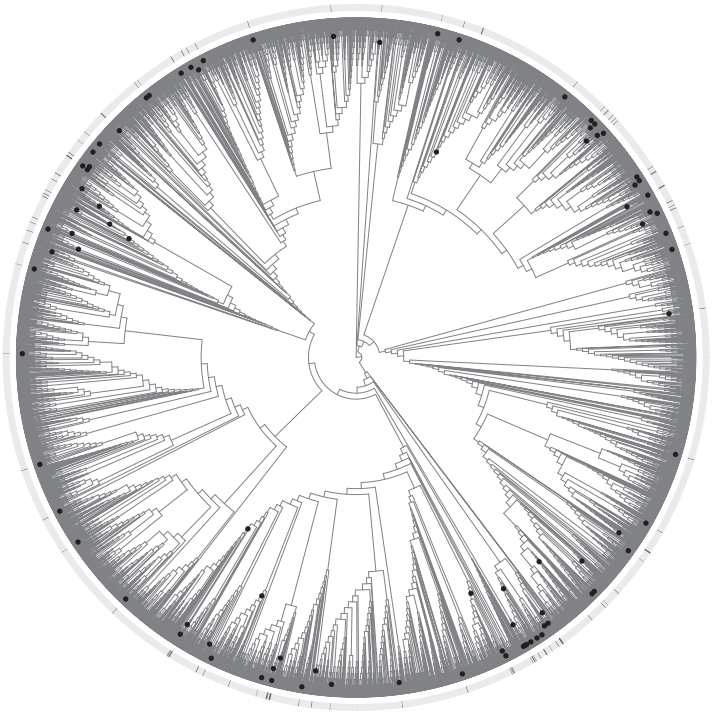
Order name (APG III)	Family name (APG III)	Species	Observed # halophytes	Observed % halophytes	*p*-value fewer	*p*-value more	Pattern
-	Salicaceae	1010	6	0.6	0.10	0.95	
Malvales	Malvaceae	4225	56	1.3	0.97	0.04	more
-	Neuradaceae	10	1	10.0	1.00	0.10	
-	Thymelaeaceae	891	3	0.3	0.02	0.99	fewer
Myrtales	Combretaceae	500	12	2.4	1.00	0.01	more
-	Lythraceae	620	21	3.4	1.00	0.00	more
-	Melastomataceae	5005	1	0.0	0.00	1.00	fewer
-	Myrtaceae	4620	47	1.0	0.50	0.56	
-	Onagraceae	656	6	0.9	0.48	0.67	
Nymphaeales	Nymphaeaceae	58	3	5.2	1.00	0.02	more
Oxalidales	Oxalidaceae	770	2	0.3	0.01	1.00	fewer
Pandanales	Pandanaceae	885	11	1.2	0.79	0.31	
Picramniales	Picramniaceae	49	1	2.0	0.91	0.40	
Piperales	Piperaceae	3615	1	0.0	0.00	1.00	fewer
-	Saururaceae	6	1	16.7	1.00	0.06	
Poales	Bromeliaceae	1770	2	0.1	0.00	1.00	fewer
-	Cyperaceae	5430	121	2.2	1.00	0.00	more
-	Flagellariaceae	4	1	25.0	1.00	0.04	more
-	Juncaceae	430	22	5.1	1.00	0.00	more
-	Poaceae	11160	335	3.0	1.00	0.00	more
-	Restionaceae	500	2	0.4	0.11	0.97	
-	Typhaceae	25	9	36.0	1.00	0.00	more
Proteales	Nelumbonaceae	2	1	50.0	1.00	0.02	more
Ranunculales	Menispermaceae	442	3	0.7	0.33	0.83	
-	Papaveraceae	760	3	0.4	0.05	0.98	fewer
-	Ranunculaceae	2525	17	0.7	0.04	0.98	fewer
Rosales	Elaeagnaceae	45	3	6.7	1.00	0.01	more
-	Moraceae	1125	7	0.6	0.11	0.94	
-	Rhamnaceae	925	6	0.6	0.16	0.91	
-	Rosaceae	2520	9	0.4	0.00	1.00	fewer
-	Ulmaceae	35	1	2.9	0.95	0.30	
Santalales	Olacaceae	57	1	1.8	0.88	0.45	
-	Santalaceae	990	3	0.3	0.01	1.00	fewer
Sapindales	Anacardiaceae	873	7	0.8	0.32	0.79	
-	Meliaceae	615	6	1.0	0.55	0.61	
-	Nitrariaceae	16	8	50.0	1.00	0.00	more
-	Rutaceae	2070	5	0.2	0.00	1.00	fewer
-	Sapindaceae	1630	2	0.1	0.00	1.00	fewer
-	Simaroubaceae	110	1	0.9	0.69	0.68	
Saxifragales	Crassulaceae	1370	2	0.1	0.00	1.00	fewer
-	Cynomoriaceae	2	1	50.0	1.00	0.02	more
Solanales	Convolvulaceae	1625	22	1.4	0.92	0.12	
-	Hydroleaceae	12	1	8.3	0.99	0.12	
-	Solanaceae	2460	41	1.7	1.00	0.00	more
Unplaced Asterid I	Boraginaceae	2755	37	1.3	0.95	0.07	
Vitales	Vitaceae	850	4	0.5	0.06	0.98	
Zingiberales	Zingiberaceae	1208	1	0.1	0.00	1.00	fewer
Zygophyllales	Zygophyllaceae	285	30	10.5	1.00	0.00	more

Figure S1: Family subtrees for the sample of 22 angiosperm families analysed. Origins of salt tolerance identified by maximum parsimony (see Methods) are marked on each family with black circles. Tips in the subtrees identified as halophytes are marked in black in the ring around the subtree.



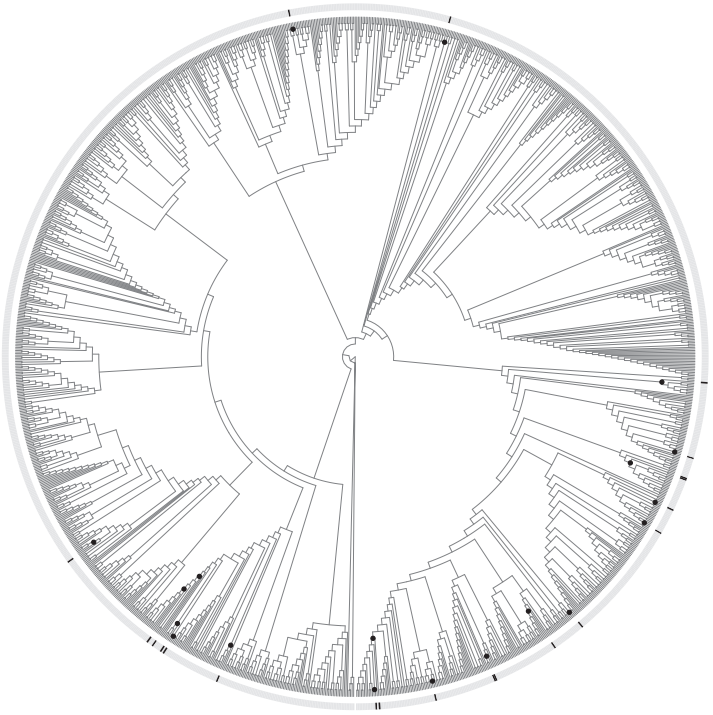


Goodeniaceae

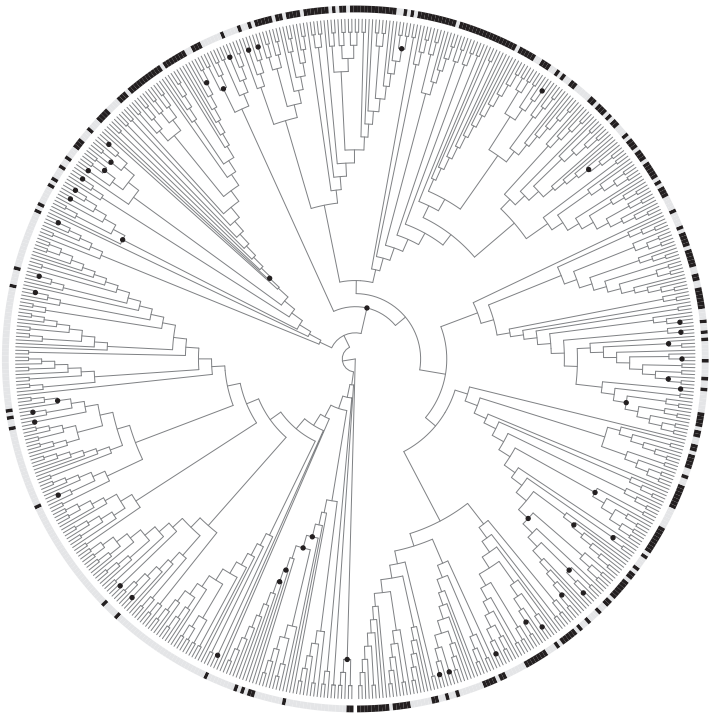


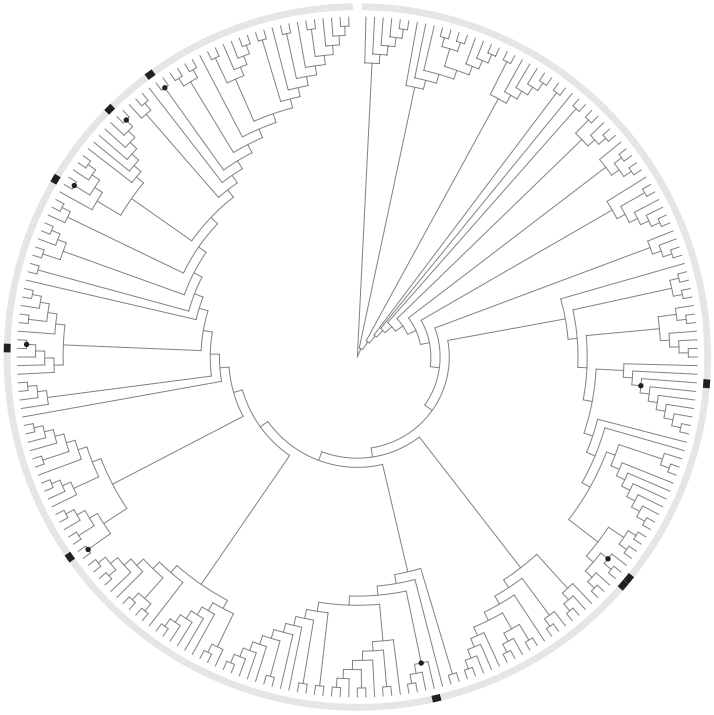
Asteraceae

Brassicaceae



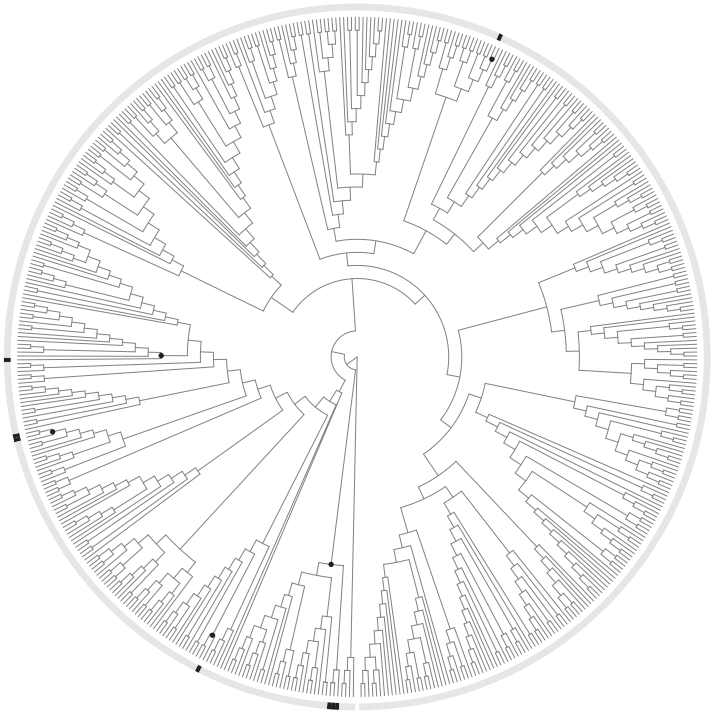
Amaranthaceae

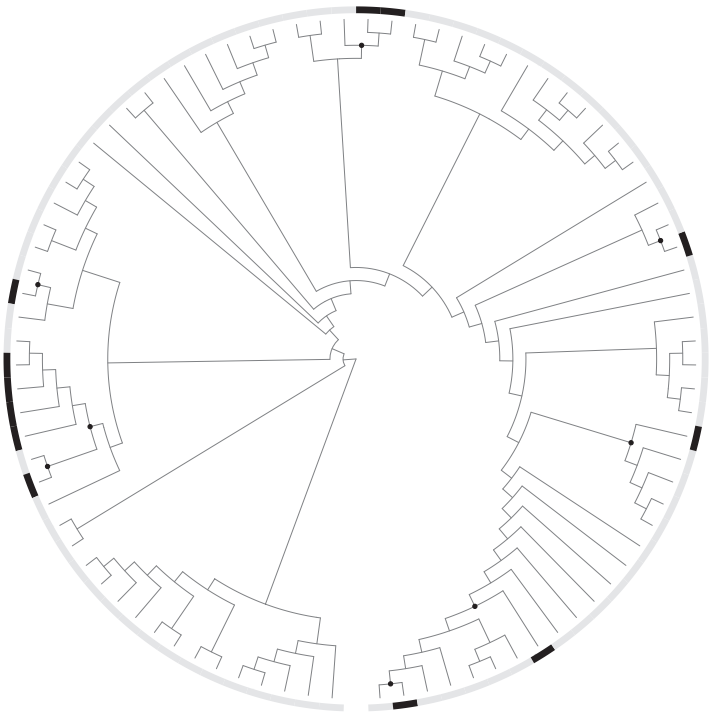


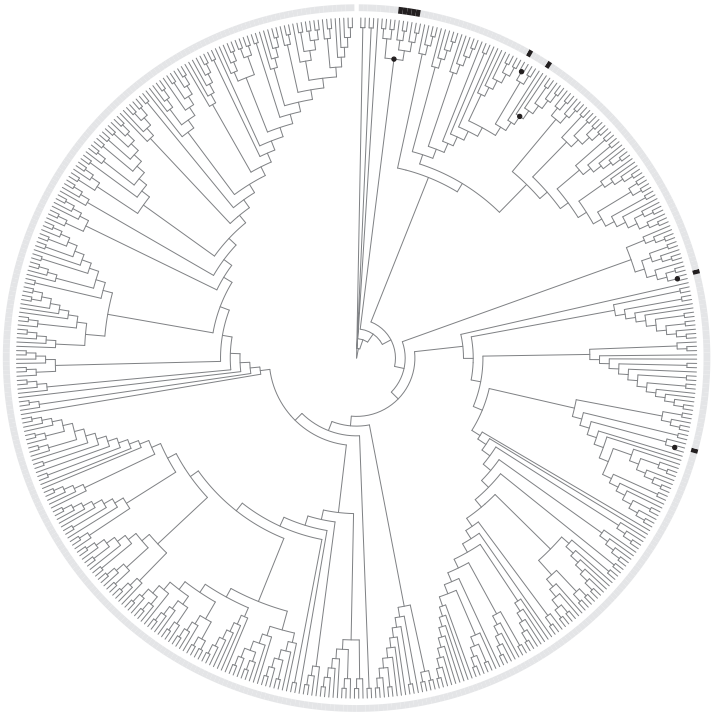


Cucurbitaceae

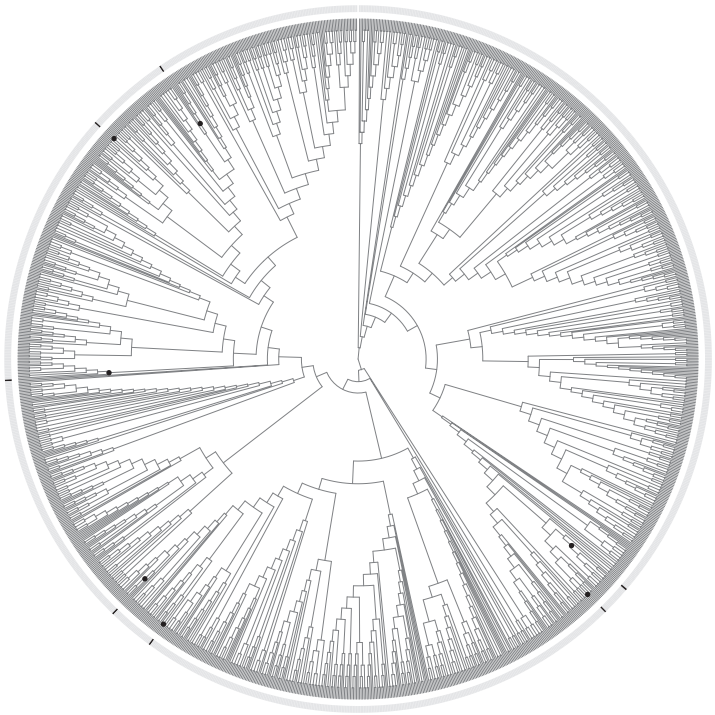
Tamaricaceae


Primulaceae

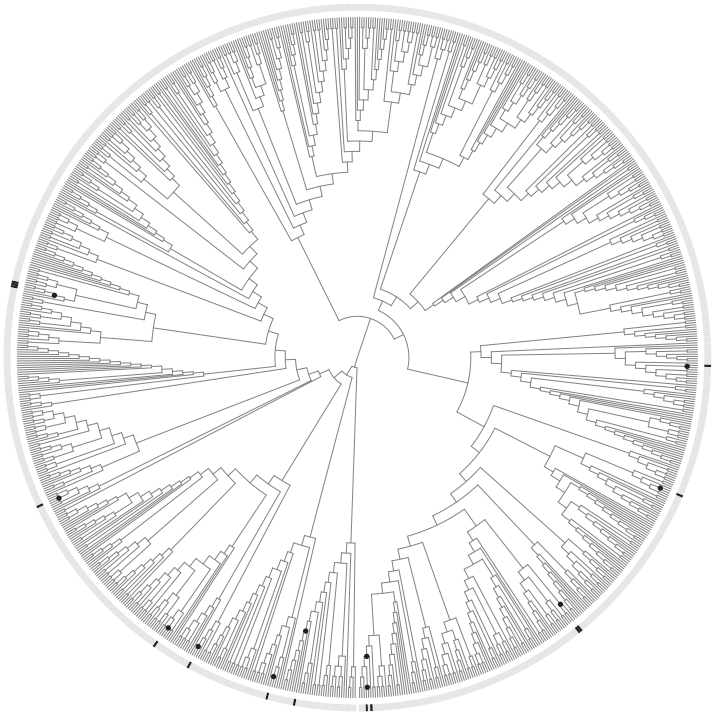

Casuarinaceae



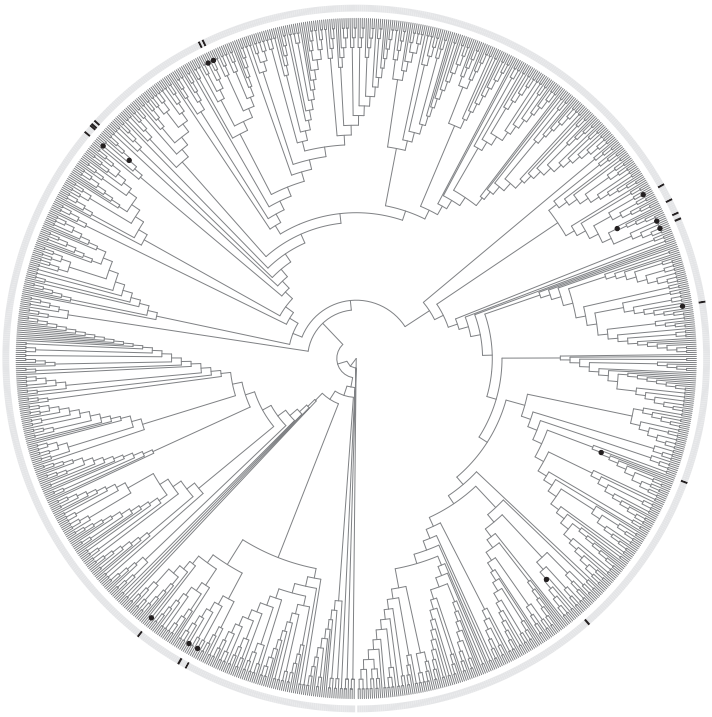
Acanthaceae



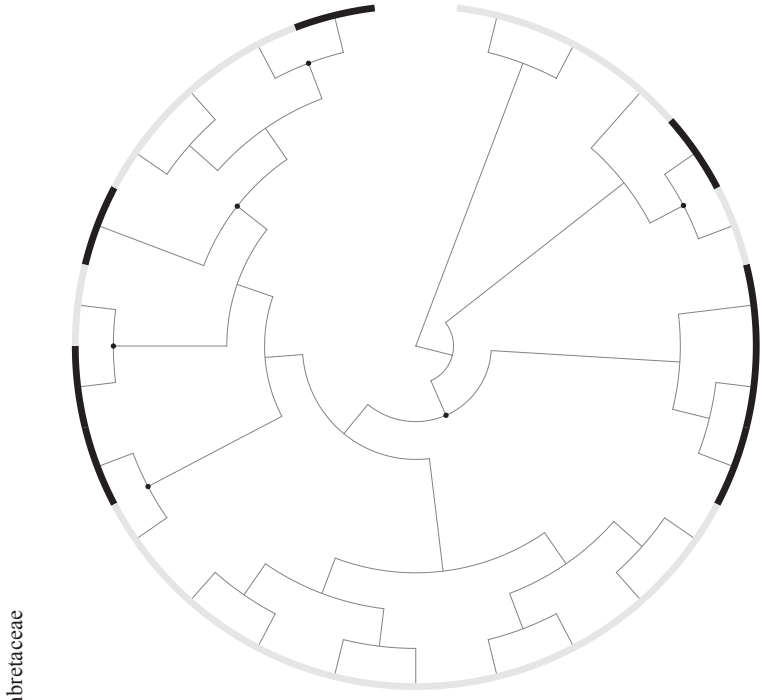
Rubiaceae



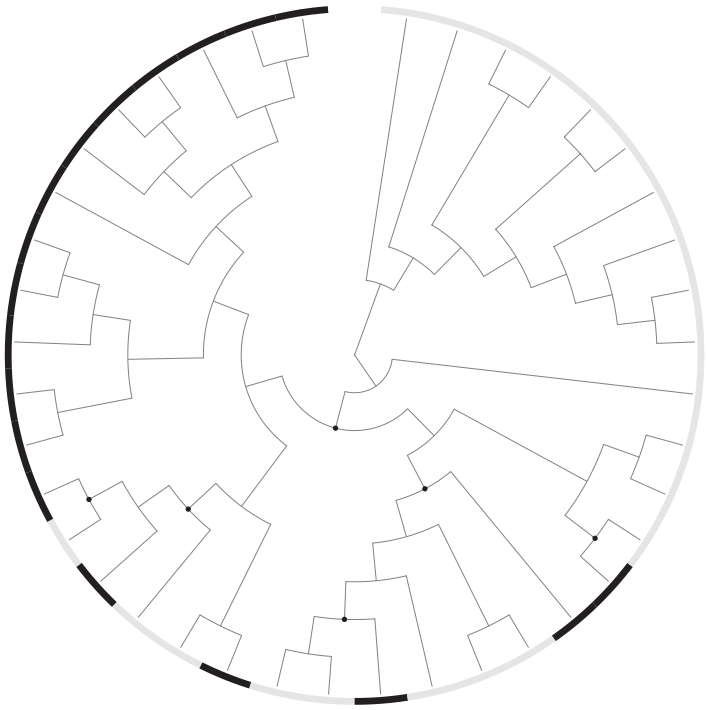
Lamiaceae



Euphorbiaceae

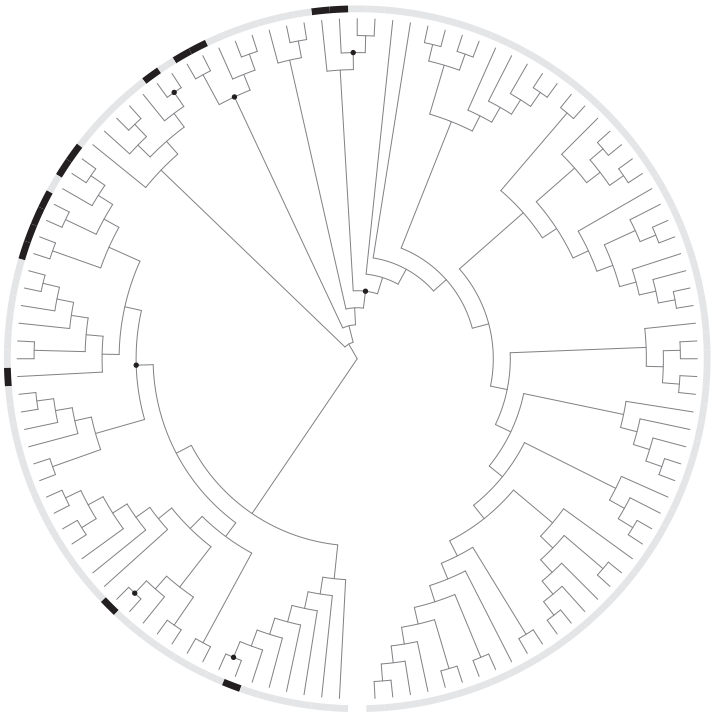


Combretaceae

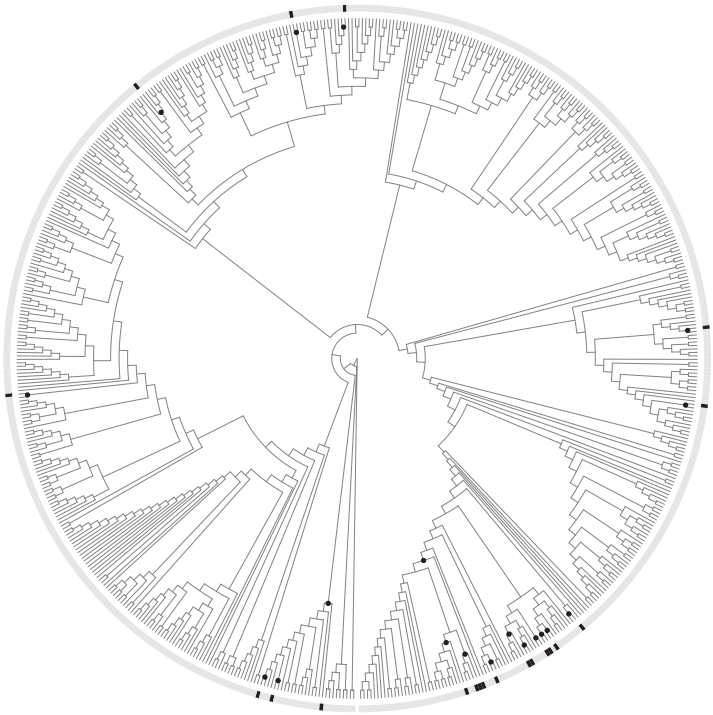


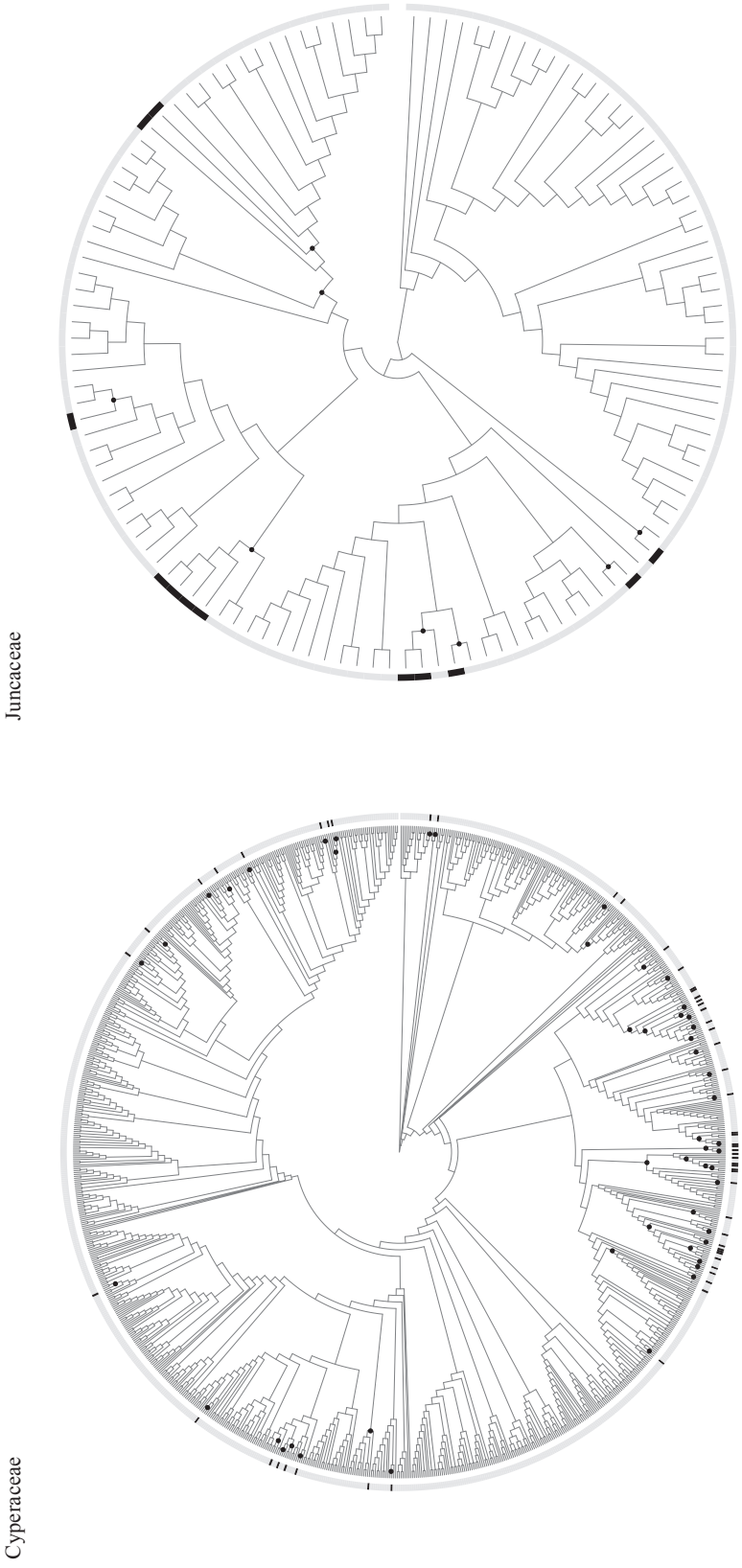
Rhizophoraceae

Lythraceae

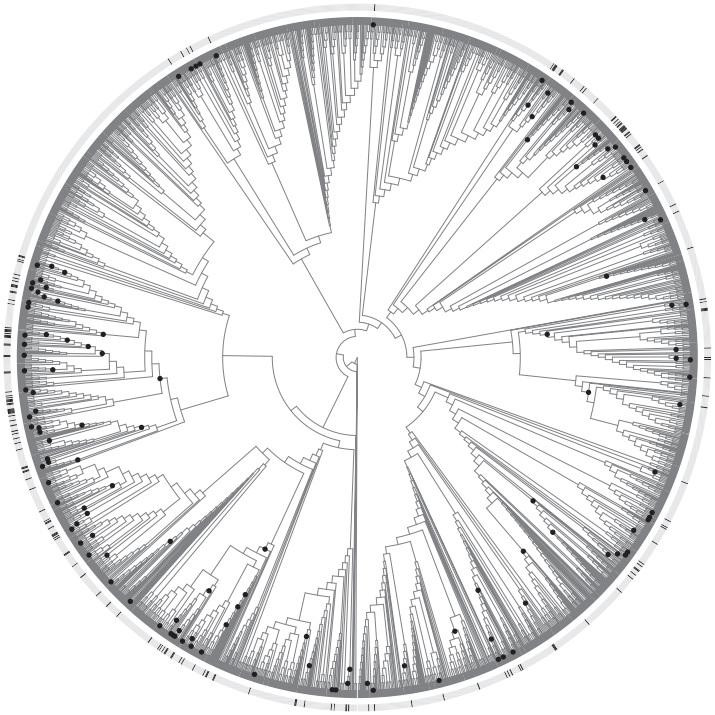


Myrtaceae

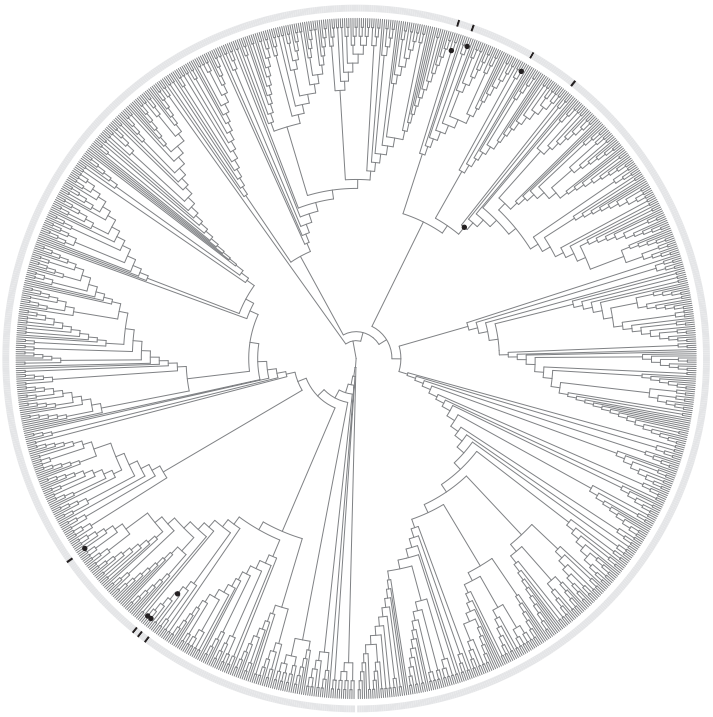



Juncaceae
Cyperaceae

Poaceae



Rosaceae



References

Dagar, J.C. & Gurbachan, S. 2007. Biodiversity of saline and waterlogged environments: Documentation, utilization and management. *NBA Scientific Bulletin Number - 9*, National Biodiversity Authority, Chennai, Tamil Nadu, India, pp. 78.

Guvensen, A., Gork, G. & Özturk, M. 2006. An overview of the halophytes in Turkey. *Sabkha ecosystems*, pp. 9–30. Springer, Netherlands.

Haston, E., Richardson, J.E., Stevens, P.F., Chase, M.W. & Harris, D.J. 2009. The Linear Angiosperm Phylogeny Group (LAPG) III: A linear sequence of the families in APG III. *Bot. J. Linn. Soc.* **161**: 128–131.

Khan, M.A. & Qaiser, M. 2006. Halophytes of Pakistan: characteristics, distribution and potential economic usages. *Sabkha ecosystems*, pp. 129–153. Springer, Netherlands.

Menzel, U. & Lieth, H. 2003. *HALOPHYTE Database V. 2.0 update.* (H. Lieth & M. Mochtchenko, eds). Kluwer, Netherlands.

Özturk, M., Guvensen, A., Sakçali, S. & Gork, G. 2008. Halophyte plant diversity in the Irano-Turanian phytogeographical region of Turkey. In: *Biosaline Agriculture and High Salinity Tolerance*, pp. 141–155. Birkhäuser, Basel.

Smith, S.A., Beaulieu, J.M., Stamatakis, A. & Donoghue, M.J. 2011. Understanding angiosperm diversification using small and large phylogenetic trees. *Am. J. Bot.* **98**: 404–414.

Stevens, P.F. 2001. Angiosperm Phylogeny Website. Version 12, July 2012. http://www.mobot.org/MOBOT/research/ APweb/

The Plant List. 2010. The Plant List. http://theplantlist.org

Zhao, K., Song, J., Feng, G., Zhao, M. & Liu, J. 2010. Species, types, distribution, and economic potential of halophytes in China. *Plant Soil* **342**: 495–509.

Chapter 3

The phylogenetic association between salt tolerance and heavy metal hyperaccumulation in angiosperms

Camile Moray, Eric W. Goolsby, Lindell Bromham

Published: Evolutionary Biology 2015. doi:10.1007/s11692-015-9355-2

Abstract

Salt tolerance and heavy metal hyperaccumulation are two rare plant abilities that are heavily studied for their potential to contribute to agricultural sustainability and phytoremediation in response to anthropogenic environmental change. Several observations suggest that it is worth investigating the link between the abilities to tolerate high levels of soil salinity or accumulate more of a particular heavy metal from the soil than most plants. Firstly, several angiosperm families are known to contain both salt tolerant plants (halophytes) and heavy metal hyperaccumulators. Secondly, some halophytes can also accumulate heavy metals. Thirdly, although salinity tolerance and heavy metal hyperaccumulation typically require many physiological or anatomical changes, both have apparently evolved many times in angiosperms and among closely related species. We test for a significant relationship between halophytes and hyperaccumulators in angiosperms using taxonomic and phylogenetic analyses. We test whether there are more angiosperm families with both halophytes and hyperaccumulators than expected by chance, and whether there are more species identified as both halophyte and hyperaccumulator than if the abilities were unconnected. We also test whether halophytes and hyperaccumulators are phylogenetically clustered among species in seven angiosperm families. We find a significant association between halophytes and hyperaccumulators among angiosperm families and that there are significantly more species identified as both halophytes and hyperaccumulators than expected. Halophytes and hyperaccumulators each show low phylogenetic clustering, suggesting these abilities can vary among closely related species. In Asteraceae, Amaranthaceae, Fabaceae, and Poaceae, halophytes and hyperaccumulators are more closely related than if the two traits evolved independently.

Introduction

The interest in understanding the ability of some plants to tolerate harsh environments has increased due to rapid anthropogenic environmental change. A large research effort has focused on identifying plants with particular traits that can tolerate and possibly mitigate the effects of these changes (Arthur et al. 2005; Bartels and Sunkar 2005; Mahajan and Tuteja 2005; Rozema and Flowers 2008; Feuillet et al. 2008). Two common environmental changes that pose challenges for land managers in both agricultural and industrialized areas are land salinization and the contamination of soils with heavy metals. For each of these problems, a group of rare, naturally occurring plants has been identified with the potential to alleviate these problems: halophytes, salt tolerant plants, and heavy metal hyperaccumulators, plants that can extract heavy metals from the soil.

As a consequence of common practices like land clearing and irrigation, approximately 7% of global land surface area is salt-affected. In particular 20-50% of irrigated agricultural land is salt-affected, which poses a significant threat to agricultural production (Munns 2005; Panta et al. 2014). Halophytes are plant species that can live in soils with salinity levels that are toxic to most plants (Glenn et al. 1999; Colmer and Flowers 2008). Halophytes are relatively rare amongst angiosperms, representing only 1-2% of flowering plant species. They have been widely studied for their potential to contribute to the expansion and sustainability of agriculture in the wake of increasing land salinization, enabling crop production on salt-affected agricultural land as well as crop production in naturally saline areas (Glenn et al. 1999; Colmer et al. 2006; Rozema and Flowers 2008; Panta et al. 2014). Halophytes have been proposed as alternative crops for food and fodder (Weber et al. 2007; El Shaer 2010; Ventura and Sagi 2013; Ventura et al. 2015), and also for their potential ability to desalinize salt-affected soils (Ravindran et al. 2007; Rabhi et al. 2010). There has also been a large research focus on how halophytes tolerate salinity, knowledge that is being used in efforts to increase salt tolerance of established crop species (Flowers and Yeo 1995; Munns et al. 2006; Colmer et al. 2006; Rozema and Flowers 2008; Tester and Langridge 2010).

Another common consequence of anthropogenic environmental change is the contamination of soils with heavy metals (such as copper, nickel, and zinc) or

metalloids (such as aluminum, arsenic, and selenium). The expansion of mining and industry has greatly increased the amount and distribution of soils contaminated with heavy metals/metalloids (Nriagu 1979), which are toxic to the vast majority of plants and pose a health risk to humans and animals. The use of some pesticides and chemical and biological fertilizers has lead to the contamination of agricultural lands with heavy metals, which can contaminate crops and fodder (Baker et al. 1994; Wuana and Okieimen 2011). Heavy metals accumulate in soils and do not dissipate over time, so it is necessary to remove or alleviate the negative effects of anthropogenic contaminates in soil and ground water.

Researchers have studied plants known as heavy metal hyperaccumulators as an alternative to chemical and physical methods of removing heavy metals from soils (Vara Prasad and de Oliveira Freitas 2003; Arthur et al. 2005). Heavy metal hyperaccumulators, referred to here as hyperaccumulators, are plant species that are able to not only tolerate but also extract large amounts of one or a few types of heavy metals from the soil into aerial tissues. Like halophytes, hyperaccumulators are rare and represent approximately 0.2% of plant species (Baker and Brooks 1989; Rascio and Navari-Izzo 2011; Cappa and Pilon-Smits 2014). Hyperaccumulators can take up hundreds or thousands of times greater concentrations of particular heavy metals/metalloids than most plants (Rascio and Navari-Izzo 2011), and so have been studied for their potential use in phytoremediation, using plants to clear or alleviate the effects of excess metals from contaminated soils (Arthur et al. 2005; Ali et al. 2013). Phytoremediation has been proposed as a cost-effective and environmentally low-impact alternative for removing or alleviating the effects of heavy metal contamination of soils. Many studies have focused on either the direct use of natural hyperaccumulators or on engineering novel hyperaccumulators for phytoremediation (Vara Prasad and de Oliveira Freitas 2003; Arthur et al. 2005; Manousaki and Kalogerakis 2011b). Hyperaccumulators have also been researched for use in phytomining, using plants to extract valuable metals/metalloids from contaminated and naturally-occurring metalliferous soils (Brooks et al. 1998; Anderson et al. 1999; Sheoran et al. 2009).

The large research effort focusing on hyperaccumulators and halophtyes has produced experimental and observational evidence that salt tolerance and heavy metal hyperaccumulation may be physiologically and evolutionarily associated. For example,

several halophytes can accumulate heavy metals, such as *Arthrocnemum macrostachyum* (Amaranthaceae) and *Tamarix smyrnensis* (Tamaricaceae) (Jordan et al. 2002; Kadukova et al. 2008; Redondo-Gómez et al. 2010; Redondo-Gómez 2013). One explanation for why some halophytes can accumulate heavy metals is that both abilities rely on similar functional mechanisms. Excess salt and heavy metals are both toxic to plants, and both salt tolerance and heavy metal hyperaccumulation are often the results of many physiological or anatomical modifications (Flowers et al. 1977; Baker and Brooks 1989). Salt and heavy metals can both induce osmotic and metabolic stresses, and halophytes and hyperaccumulators may use similar mechanisms to combat these stresses (Flowers et al. 1977; Baker and Brooks 1989; Thomas et al. 1998; Przymusiński et al. 2004). For example, one effect of toxic levels of metals and salts within plants is the increased production of reactive oxygen species (ROS; Briat and Lebrun 1999; Bose et al. 2014), which unchecked can lead to cell damage and plant death. Some halophytes and hyperaccumulators use the same mechanisms for dealing with ROS, including the production of compatible solutes, which act as osmoprotectants (Schat et al. 1997; Glenn et al. 1999; Sharma and Dietz 2006; Munns and Tester 2008; Lefèvre et al. 2009). In some cases, halophytes and hyperaccumulators produce the same osmoprotectants, like proline (Stewart and Lee 1974; Flowers et al. 1977; Schat et al. 1997; Sharma and Dietz 2006). Some halophytes and hyperaccumulators are also known to use shedding to deal with excess toxins, pushing salts and metals into leaves or other aerial tissues and then shedding them to remove toxins (Albert 1975; Boyd 2004). Specific anatomical adaptations may also allow for some species to tolerate and remove heavy metals and salts. For example, studies have shown that specialized salt glands, which extrude excess salt out of the plant body, are also able to extrude multiple types of heavy metals/metalloids (Jordan et al. 2002; Kadukova et al. 2008; Manousaki and Kalogerakis 2011b).

In addition to the observation that some species are identified as both halophytes and hyperaccumulators, there also appears to be a broader taxonomic and evolutionary association between halophytes and hyperaccumulators among plant families. Although halophytes and hyperaccumulators are rare, they are found in a diverse range of angiosperm families. Several angiosperm families, including Asteraceae, Euphorbiaceae and Brassicaceae contain both halophyte and hyperaccumulator species (Flowers et al. 1977; Vara Prasad and de Oliveira Freitas 2003; Menzel and Lieth 2003; Rascio and Navari-Izzo 2011). One possible explanation for the co-occurrence of halophytes and

hyperaccumulators in these families is that some feature of these lineages may make the evolution of salt tolerance, heavy metal hyperaccumulation, or both, more likely.

By comparing phylogenetic studies, it also appears that salt tolerance and heavy metal hyperaccumulation show some similar evolutionary patterns. Although salt tolerance and heavy metal hyperaccumulation often involve multiple physiological or anatomical mechanisms, phylogenetic and taxonomic evidence suggests that there have been many independent evolutionary origins of both salt tolerance (Flowers et al. 2010; Bennett et al. 2013; Saslis-Lagoudakis et al. 2014) and heavy metal hyperaccumulation (Cappa and Pilon-Smits 2014). Phylogenetic analyses suggest that salt tolerance has evolved many times among species within several families (Bennett et al. 2013; Moray et al. 2015). And it has also been suggested that heavy metal hyperaccumulation has evolved multiple times within some families and genera (Krämer 2010; Cecchi et al. 2010; Cappa and Pilon-Smits 2014). These observations suggest that salt tolerance and heavy metal hyperaccumulation may both evolve more often in some taxonomic groups than expected considering their rarity amongst species.

The observed association between salt tolerance and heavy metal hyperaccumulation creates an opportunity to explore whether having a particular tolerance to one environmental stress is associated with the ability to tolerate other types of stresses. One way to establish whether salt tolerance and heavy metal hyperaccumulation are associated is to use taxonomic information to find out which groups (e.g., angiosperm families) contain both halophytes and hyperaccumulators and to identify which species are identified as both a halophyte and a hyperaccumulator. But knowing whether halophytes and hyperaccumulators are related taxonomically does not fully answer the question of whether the two abilities are closely related in an evolutionary context. Using a phylogenetic comparative approach we can test not only whether salt tolerance and heavy metal hyperaccumulation are found in the same broad groups or occur in some of the same species, but also whether halophytes and hyperaccumulators are closely related among species. For example, if halophytes and hyperaccumulators are often found in closely related lineages, this could mean that within families, some lineages are more likely to produce both types of species, and others are more likely to produce none. Understanding the evolutionary relatedness between these traits could lead to the identification of factors that support the ability to tolerate multiple harsh conditions, which could contribute to the production of novel varieties of tolerant and

multi-tolerant plants for practical use (Manousaki and Kalogerakis 2011a; Hamed et al. 2013; Anjum et al. 2014; Lutts and Lefevre 2015). In this study we take an important first step towards achieving these goals by establishing whether there is a significant taxonomic association and phylogenetic relationship between halophytes and hyperaccumulators in the angiosperms.

Using lists of species identified in published sources as halophytes and hyperaccumulators, we first investigate the broader relationship between salt tolerance and heavy metal hyperaccumulation in angiosperms. We begin by asking whether there are more angiosperm families that have both halophytes and hyperaccumulators than expected. Then, using the phylogenies of seven angiosperm families, we test whether salt tolerance and heavy metal hyperaccumulation have a tendency to occur in closely related lineages by testing whether halophyte and hyperaccumulator species are more closely related than predicted by a model where each ability evolves independently. We also identify multi-tolerant species (species that are identified as both a halophyte and hyperaccumulator), and investigate whether there are more multi-tolerant species among angiosperms than expected given the rarity of both tolerances.

Methods

Taxonomic data

We compiled lists of angiosperm species reported to be hyperaccumulators or halophytes. Both heavy metal hyperaccumulation and salt tolerance can be considered on continuous scales (for example, some species can tolerate higher concentrations of salt than others), but continuous measures of tolerance/accumulation are available for relatively few species. Since we wanted to analyze the relationship between all species known to tolerate salt or hyperaccumulate heavy metals, we had to treat each ability as a binary character. Categorizing species as able to hyperaccumulate heavy metals or not, or as salt tolerant or not, allowed us to include a wider range of published sources, so that we could include species identified by both observational and experimental evidence. We included species identified as a halophyte or hyperaccumulator in published field studies and surveys, as well as halophytes and hyperaccumulators

identified in laboratory and greenhouse experiments. We analyzed the relationship between halophytes and hyperaccumulators at the species level, so we considered a species to have the propensity to tolerate salinity or hyperaccumulate heavy metals/metalloids if one or more variety or subspecies was identified as a halophyte or a hyperaccumulator in the literature.

Heavy metal hyperaccumulator list

To create a list of hyperaccumulators, we searched the Web of Science (2015) with the term "hyperaccum*" to find published reports of angiosperm species with the ability to hyperaccumulate metals (see Supplemental Material for list of references). We included species that the authors reported as hyperaccumulators. We did not restrict our list to species able to tolerate or accumulate a specific amount of metal since this information available for relatively few species and because measures of tolerance and accumulation can vary in different experimental conditions (Goolsby and Mason 2015). The resulting list had 593 species. We also added 54 species from a published list (Cappa and Pilon-Smits 2014). Because hyperaccumulators may be able to tolerate and take up one or a few particular heavy metals/metalloids, we recorded the elements accumulated by each species where available. However, because we treat hyperaccumulation as a binary trait, we did not take into account metal specificity in our analysis.

Halophyte list

We used a list of halophytes from Moray et al. (2015). This list included about 2600 taxa reported to grow in saline habitats (Menzel and Lieth 2003) as well as taxa from five additional published halophyte lists (Guvensen et al. 2006; Khan and Qaiser 2006; Dagar and Gurbachan 2007; Öztürk et al. 2008; Zhao et al. 2011). The complete list contained 3468 taxa reported to be salt tolerant (including infraspecific taxa).

Association between halophytes and heavy metal hyperaccumulators

Family-level taxonomic association

In order to identify species that are reported as both hyperaccumulators and halophytes, we needed to be sure that both lists followed a consistent taxonomy. We used the

function ‘TPL’ in the R package *taxonstand* (Cayuela et al. 2012) to search for accepted names of each taxon based on The Plant List (2010) taxonomy. This search resulted in a list of 531 accepted hyperaccumulator species. After removing infraspecific epithets and comparing the list of halophytes from the literature to The Plant List (2010), we identified 2934 accepted halophyte species.

Our first aim was to investigate the observation that several angiosperm families are known to contain both halophytes and hyperaccumulators. We tested whether there were more families containing both halophytes and hyperaccumulators than expected if the two were distributed randomly with respect to each other, accounting for the total number of species in each family and the observed proportions of halophytes and hyperaccumulators among angiosperms. Using the lists of accepted hyperaccumulators and halophytes, we first identified which families had at least one hyperaccumulator and one halophyte based on The Plant List (2010) taxonomy. We included 411 angiosperm families, by checking the 413 families identified by the Linear Angiosperm Phylogeny Group III (Haston et al. 2009) against those listed on the APG III website (Stevens 2001). Two families, Aristolochiaceae and Lactoridaceae are considered one family by the APG III website and Buxaceae and Haptanthaceae are also considered synonymous (Stevens 2001). Here we considered these families as synonyms, reducing the number of angiosperm families included in this analysis from 413 to 411. We also recognized Ripogonaceae (Haston et al. 2009) as an alternative spelling of Rhipogonaceae (APG 2009). We collected an estimate of the number of species in each family, by taking the mean of the species estimates listed for each family on the APG III website (Stevens 2001). We also estimated the observed proportion of species identified as either a halophyte or hyperaccumulator among the total of 276,000 angiosperm species across all families. We compared the observed number of families with both one or more halophytes and hyperaccumulators to a Poisson binomial distribution, using the ‘ppoibin’ function in the R package *poibin* (Hong 2013), parameterized by the observed number of families identified as having at least one halophyte and one hyperaccumulator, and the probability of each angiosperm family having both a hyperaccumulator and halophyte given the observed proportions of each ability among all angiosperm species and the estimated number of species in each family.

Frequency of multi-tolerant species

Next we asked whether salt tolerance and heavy metal hyperaccumulation occurred in the same species more often than expected given the rarity of both abilities. We tested if there were more species that were included in both the lists of accepted halophyte and hyperaccumulator species than expected by chance. Using the estimates of total angiosperm species calculated in the family-level taxonomic analysis, we calculated the observed frequencies of halophytes, hyperaccumulators, multi-tolerant and non-tolerant species among angiosperm species, and the expected probabilities of each species only being a halophyte, only being a hyperaccumulator, being a multi-tolerant species, or not having either ability. We then used a X^2 test for given probabilities to ask if the observed frequency of multi-tolerant species was significantly greater than predicted by the expected probabilities.

Phylogenies

We also aimed to assess the phylogenetic relatedness between halophytes and hyperaccumulator species within families. We used a published phylogeny of over 56,000 angiosperm taxa (Smith et al. 2011) to extract species-level trees for a number of angiosperm families. In order to select informative examples for analysis, we needed to target families that had enough halophytes and hyperaccumulators to allow us to test the phylogenetic relationship between the two. We first identified family clades in the phylogeny that had six or more terminal taxa (tips) in the phylogeny matching species on the halophyte list and six or more tips matching species on the hyperaccumulator lists. We then created a family-level phylogeny for each of the seven families that met these criteria. If all tips associated with a family were monophyletic in the Smith et al. (2011) angiosperm phylogeny, we extracted all taxa associated with the family according to GenBank taxonomy. For non-monophyletic families, we only included species that fell within the main clade of the family (see Supplementary Material for list of excluded tips). In some cases, we also removed a small number of tips from the family clade that were not associated with the target family (see Supplementary Material for details). We then removed tips from the family trees that were not identified by a standardized genus-species epithet, as we could not confidently match

them to the lists of halophytes and hyperaccumulators. Specifically we excluded any tips that included the taxonomic epithets “af”, “aff”, “cf”, or “sp”. We also removed any tips representing hybrid taxa, by removing tips that included one genus and two specific epithets separated by “x” or that included the word “hybrid”. We randomly resolved polytomies in the family trees using the ‘multi2di’ command in the R package *ape* (Paradis et al. 2004) since polytomies can not be analyzed using the phylogenetic metrics used in this study.

Because we analyze the relationship between halophytes and hyperaccumulators at the species level, we relabeled the tip labels of infraspecific taxa in the family trees to the species name. Removing infraspecific epithets from tip labels sometimes resulted in multiple tips representing the same species. For each set of duplicate tips, we determined which tip had the most reliable position in the tree by choosing the tip with the most data in the published alignment that was also grouped with conspecifics and congenerics. The remaining duplicates were removed from the tree. Since the Smith et al. (2011) angiosperm phylogeny does not follow The Plant List (2010) taxonomy, tips were identified as a halophyte or hyperaccumulator in the phylogenetic analyses if they matched either the accepted name identified on The Plant List (2010) or the name in the halophyte/hyperaccumulator lists, which were the names presented in the surveyed publications.

Species-level phylogenetic association

Our next aim was to assess the species-level phylogenetic association between heavy metal hyperaccumulators and halophytes in different angiosperm families. The functional and taxonomic similarities between salt tolerance and heavy metal hyperaccumulation, including the observation that some species are both halophytes and hyperaccumulators, leads to the prediction that hyperaccumulator species might be quite closely related to halophytes within families. To interpret the relatedness between halophytes and hyperaccumulators, we also needed to understand the relatedness among halophytes and among hyperaccumulators. Salt tolerance has been shown to be remarkably labile in some angiosperm families, with a relatively large number of inferred independent evolutionary origins (Bennett et al. 2013; Moray et al. 2015). Several studies suggest that heavy metal hyperaccumulation has also evolved many times independently within angiosperm families (Rascio and Navari-Izzo 2011; Cappa

and Pilon-Smits 2014), but the species level phylogenetic relationships have not been formally analyzed. To distinguish patterns particular to hyperaccumulators or halophytes from the relationship between the two groups, we measured phylogenetic relatedness 1) among hyperaccumulators, 2) among halophytes, and 3) between hyperaccumulator and halophyte species in a sample of angiosperm families.

Phylogenetic relatedness among halophytes and hyperaccumulators

To measure phylogenetic relatedness among halophytes and among hyperaccumulators in each angiosperm family chosen for analysis, we measured the mean nearest taxon distance (MNTD) for each group using the function 'mntd' in the R package *picante* (Kembel et al. 2010). MNTD (derived from nearest taxon index, NTI, Webb et al. 2002) measures the mean phylogenetic distance between each taxon in a group to the closest relative within that group. A smaller MNTD indicates that the taxa in a group are more phylogenetically related than taxa with a larger MNTD. To assess the significance of the observed MNTD for each group (halophytes or hyperaccumulators) in each family, we compared the observed values to two null models. We first compared the observed MNTD to the MNTD values from 1000 random distributions, generated by randomly assigning tips in each family tree as either halophyte, hyperaccumulator or neither, constraining the total number of halophytes and hyperaccumulators in each randomization to the observed number in each family tree. The *p*-value for each family was generated by the proportion of random comparisons with a MNTD smaller than the observed. *P*-values less than or equal to 0.05 indicated that the observed MNTD was significantly smaller than 95% of the random samples.

We then compared the observed MNTD for each group in each family to a Brownian motion (BM) model. We simulated the evolution of two independent traits, which we labeled salt tolerance and heavy metal hyperaccumulation, as continuous traits using a Brownian Motion (BM) model of evolution (Felsenstein 2005; Fritz and Purvis 2010). We then converted each continuous trait to a binary one using an appropriate threshold, ensuring that the resulting number of halophyte or hyperaccumulator tips in each simulated dataset was equal to the observed numbers in each family. We repeated this process 1000 times for salt tolerance and 1000 times for heavy metal hyperaccumulation, and then measured the MNTD for each simulation. The *p*-values representing phylogenetic relatedness among halophytes and among hyperaccumulators

for each family was generated by the proportion of BM comparisons with a MNTD smaller than the observed. *P*-values less than or equal to 0.05 indicated that the observed MNTD was significantly smaller than in 95% of the BM simulations, suggesting that the species with that ability were more closely related on the phylogeny than expected under BM.

Phylogenetic relatedness between halophytes and hyperaccumulators

We then measured the phylogenetic distance between halophyte and hyperaccumulator species in each family phylogeny to ask whether, on average, halophytes and hyperaccumulators were more closely related to each other than expected. To do this, we used the between-community mean nearest taxon distance (BMNTD), a beta diversity metric performed using the 'comdistnt' function in the R package *picante* (Kembel et al. 2010). This function measures the phylogenetic distance between each taxon in one group (e.g., halophytes) and its closest relative in a second group (e.g., hyperaccumulators), and then calculates the mean of these distances. The more closely related hyperaccumulators and halophytes are to each other, the smaller the BMNTD statistic.

Since we wanted to know about the evolutionary association between salt tolerance and heavy metal hyperaccumulation, we compared the observed BMNTD to the expected pattern under a model where salt tolerance and heavy metal hyperaccumulation evolved independently under Brownian motion. Using the simulations described above, we measured the BMNTD between one simulated halophyte distribution and one simulated hyperaccumulator distribution for each of the 1000 simulations generated for each ability. The *p*-value was the proportion of simulated Brownian motion comparisons with a BMNTD smaller than the BMNTD of the observed distribution. *P*-values less than or equal to 0.05 indicated that the observed BMNTD was significantly smaller than 95% of the simulations, suggesting that halophytes and hyperaccumulators were more closely related on the phylogeny than expected if salt tolerance and heavy metal hyperaccumulation evolved independently under BM.

Results

Association between halophytes and heavy metal hyperaccumulators

Family-level taxonomic association

Of the 411 angiosperm families included in the analysis, we identified 82 families that have at least one hyperaccumulator and 149 that had at least one halophyte species (see Table S1). There were 62 families that contained both halophytes and hyperaccumulators, which is significantly more than expected by a Poisson binomial distribution parameterized by the observed proportion of halophytes and hyperaccumulators and the size of each family ($p < 0.001$). A family-level phylogenetic plot highlighting the families with halophytes and hyperaccumulators is presented in Figure 1.

Frequency of multi-tolerant species

We found that 60 species appeared on both the list of known halophyte species and the list of known hyperaccumulator species (see Table S2 for list of multi-tolerant species), representing 21 families in 15 orders (Table 1). The number of multi-tolerant species was much higher than expected based on the proportion of known halophytes and hyperaccumulators among angiosperm species (X^2 test for given probabilities, $p < 0.001$).

Phylogenetic relatedness among halophytes and hyperaccumulators

In six of the seven families, halophytes showed low phylogenetic relatedness: halophytes were less related than expected under Brownian motion (Table 2), but more closely related than a random distribution. Similarly, heavy metal hyperaccumulators were less clustered than expected under a Brownian motion model in four of the seven families. And in another four families, hyperaccumulators were more closely related, or clustered, than expected under a random distribution. These results indicate that the phylogenetic distribution of halophytes and hyperaccumularos both have low

phylogenetic relatedness in several families, but are often distinguishable from a random distribution.

Phylogenetic relatedness between halophytes and hyperaccumulators

In four of the seven families (Asteraceae, Amaranthaceae, Fabaceae, Poaceae) examined using species-level phylogenies, hyperaccumulators and halophytes were more closely related than if the two abilities had evolved independently of each other under Brownian motion (Table 3). In the remaining three families (Brassicaceae, Euphorbiaceae, Phyllanthaceae), the phylogenetic distance between halophytes and hyperaccumulators was indistinguishable from a model where both abilities evolved independently.

Discussion

In this study, we investigated whether there is a significant taxonomic and phylogenetic relationship between the ability to tolerate soil salinity and to hyperaccumulate heavy metals from the soil. Using broad scale taxonomic approaches, we find that salt tolerance and heavy metal hyperaccumulation are significantly associated among angiosperm families, as there are more angiosperm families containing both halophytes and hyperaccumulators than expected. We also find that the there are significantly more species identified as both halophytes and hyperaccumulators than expected, given the rarity of both abilities.

These findings provide evidence that there is a significant (non-random) association between salt tolerance and heavy metal hyperaccumulation in angiosperms. Furthermore, in four of the seven families that we analyzed, halophytes and hyperaccumulator species are more closely related to each other than predicted by a model of independent trait evolution, suggesting that salt tolerance and heavy metal hyperaccumulation are non-randomly distributed across lineages in these families.

The observation that more angiosperm families contain both halophytes and hyperaccumulators than expected suggests that some families are more likely to produce both halophytes and hyperaccumulators than others. By inspecting the data (Table S1),

this pattern does not seem to be driven by the prevalence of multi-tolerant species that can both tolerate salinity and hyperaccumulate heavy metals. One explanation for why some families produce both types of species is that these families have underlying "enabling traits" (Edwards and Donoghue 2013) that may support the ability to tolerate excess salinity or hyperaccumulate heavy metals. For example, exposure to excess salinity and heavy metals both induce osmotic stress, so it could be that halophytes and hyperaccumulators evolve more often in families with pre-existing adaptations to other environmental stresses that induce osmotic stress such as drought or aridity. In support of this idea, there is evidence that salt tolerance evolves more often in lineages that use C_4 photosynthesis (Sage 2004; Bromham and Bennett 2014). C_4 photosynthesis is associated with increased water use efficiency in arid environments (Sage 2004), so it could be that C_4 plants can more readily tolerate osmotic stress from excess salinity (Bromham and Bennett 2014). Similarly, heavy metal hyperaccumulation may also be associated with drought tolerance (Proctor 1999; Hughes et al. 2001; Anacker 2014). Many hyperaccumulators are endemic to serpentine habitats, which are often arid and experience drought conditions (Proctor 1999; Hughes et al. 2001; Anacker 2014), and experimental evidence suggests that a plant's response to drought and heavy metals are similar (de Silva et al. 2012). Some evidence also suggests that accumulated heavy metals may even play a role in increasing drought tolerance (Bhatia et al. 2005).

By compiling and comparing lists of halophytes and hyperaccumulators, we have identified 60 species from a diverse range of angiosperm groups that are able to both tolerate salt and hyperaccumulate heavy metals. Based on the proportion of known halophytes and hyperaccumulators among angiosperms, and assuming that the two abilities are taxonomically independent, we would predict only a few angiosperm species to have both abilities. Therefore, there are many more multi-tolerant species than expected if there were no link between salt tolerance and heavy metal accumulation. The identification of significantly more multi-tolerant species than expected provides further evidence that physiological mechanisms can allow species to both tolerate salinity and hyperaccumulate heavy metals (Anjum et al. 2014). Previous work on the use of halophytes for phytoremediation of heavy metals has focused on highly salt tolerant halophytes with specialized anatomical salt glands that can also excrete heavy metals (Kadukova et al. 2008). But the 60 multi-tolerant species we identify come from a broad range of families and orders. Not all of the angiosperm orders identified are known to have species with salt glands (Flowers et al. 2010), which

suggests that the ability to tolerate salt and hyperaccumulate heavy metals is not only determined by the presence of these specialized anatomical features. We hope this list of species (Table S2) will be useful in future studies into common mechanisms involved in salt tolerance and heavy metal hyperaccumulation as well as in research identifying species for phytoremediation.

We also find that in some families, halophyte and hyperaccumulator species are significantly more closely related phylogenetically than expected if the two abilities evolved independently under Brownian motion. This pattern might indicate that in these families salt tolerance and heavy metal hyperaccumulation are more likely to evolve in the same lineages. If this is true, these families might be good targets for future studies on the evolution of multiple stress tolerance and the identification and development of halophytic-hyperaccumulator species for use in phytoremediation. However, we only find that halophytes and hyperaccumulators are significantly related in a few families, suggesting that the relationship between salt tolerance and heavy metal hyperaccumulation may not be consistent among angiosperm families. Our results may be influenced by incomplete data on halophytic and hyperaccumulating species as well as incomplete phylogenetic sampling. It is likely that more halophytes and hyperaccumulators will be identified in future, which could change our understanding of how these abilities are related. The phylogenetic tree of angiosperms used in this study (Smith et al. 2011) represents 10% of angiosperm species, so complete sampling of angiosperm taxa would further clarify our understanding of the relationship between halophytes and hyperaccumulators.

Our results for phylogenetic relatedness among halophytes and hyperaccumulators suggest that both abilities have low phylogenetic relatedness. Inspection of the family phylogenies (Figure S2) suggests that both halophytes and hyperaccumulators are scattered across the phylogenies, rather than being clustered into a few clades containing many tolerant species, supporting previous findings that both traits may be labile amongst angiosperm species (Bert et al. 2003; Greenwood and MacFarlane 2009; Cecchi et al. 2010; Bennett et al. 2013; Cappa and Pilon-Smits 2014; Moray et al. 2015). One explanation for this pattern is that both traits can evolve over short time scales. For example, the amount of salt that halophytes can tolerate and the amount of metal hyperaccumulators can retain can vary not only between closely related species (Bert et al. 2003; Greenwood and MacFarlane 2009; Cecchi et al. 2010; Rozema et al.

2014), but even between populations of the same species (Antonovics et al. 1971; Wu et al. 1975; Reeves et al. 2001). Furthermore, some of the mechanisms for salt tolerance and heavy metal hyperaccumulation involve the regulation or alteration of existing functions rather than the development of novel structures like salt glands (Flowers et al. 1977; Hanikenne and Nouet 2011). If regulatory changes are more labile than anatomical features or are more likely to occur in some lineages, this could contribute to the repeated evolution of these abilities.

In this study, we have analyzed salt tolerance and the ability to hyperaccumulate heavy metals as binary characters in order to allow us to include the maximum number of species and look at broad patterns across angiosperms. If continuous measures of tolerance were available for more species, it would permit a closer examination of the links between these tolerances, and may have practical benefits. For example, identifying species that have very high salt tolerance and can also accumulate multiple types of metals may be most useful for phytoremediation of contaminated salt marshes/lakes (Redondo-Gómez et al. 2010).

Conclusions

A large research effort has focused on the use of halophytes and heavy metal hyperaccumulators for practical use. Several observations have highlighted the physiological and taxonomic association between salt tolerance and heavy metal hyperaccumulation as well as the similarities in their patterns of evolution. We confirm that there is a significant taxonomic association between salt tolerance and heavy metal hyperaccumulation in angiosperms: significantly more angiosperm families contain both halophytes and hyperaccumulators than expected and there is a significantly large number of angiosperm species that can both tolerate salinity and hyperaccumulate heavy metals. Both tolerances are scattered across the phylogenies of several families and have low phylogenetic relatedness, suggesting that salt tolerance and heavy metal hyperaccumulation may vary among closely related species. Halophytes and hyperaccumulators are significantly closely related to each other in some families, but we do not find evidence that this pattern is consistent across angiosperm families. We hope that the identification of families with a significant association between salt

tolerance and heavy metal hyperaccumulation and the identification of a large and diverse set of multi-tolerant species will contribute to future advances in phytoremediation and agricultural sustainability.

Acknowledgements

We would like to thank Marcel Cardillo, Xia Hua, and Haris Saslis-Lagoudakis for providing helpful advice and feedback on the methods used in this analysis.

Page intentionally left blank

Table 1: Angiosperm families that include species able to tolerate salinity and hyperaccumulate heavy metals. Family size is the mean of the number of estimated species from each family (Stevens 2001), halos is the number of known halophytes, hypers is the number of known heavy metal hyperaccumulators, and multi are the species that are identified to both tolerate salinity and hyperaccumulate heavy metals. A complete list of the multi-tolerant species identified is presented in Table S2.

Order	Family	Family size	Halos	Hypers	Multi
Alismatales	Araceae	4759	8	4	2
Asparagales	Iridaceae	2025	10	5	2
Asterales	Asteraceae	23600	275	84	9
Brassicales	Brassicaceae	3710	40	92	3
Caryophyllales	Aizoaceae	2035	46	2	2
-	Amaranthaceae	2275	508	11	7
-	Plumbaginaceae	836	62	1	1
-	Polygonaceae	1110	41	7	1
Commelinales	Pontederiaceae	33	3	1	1
Fabales	Fabaceae	19500	252	27	4
Gentianales	Apocynaceae	4555	44	2	1
Lamiales	Lamiaceae	7173	31	12	2
-	Plantaginaceae	1900	35	2	1
Malpighiales	Euphorbiaceae	5735	43	37	3
Malvales	Malvaceae	4225	56	8	2
Myrtales	Lythraceae	620	23	2	1
Poales	Cyperaceae	5430	124	8	1
-	Poaceae	11160	345	29	14
Solanales	Convolvulaceae	1625	22	5	1
-	Solanaceae	2460	42	2	1
Zygophyllales	Zygophyllaceae	285	30	1	1

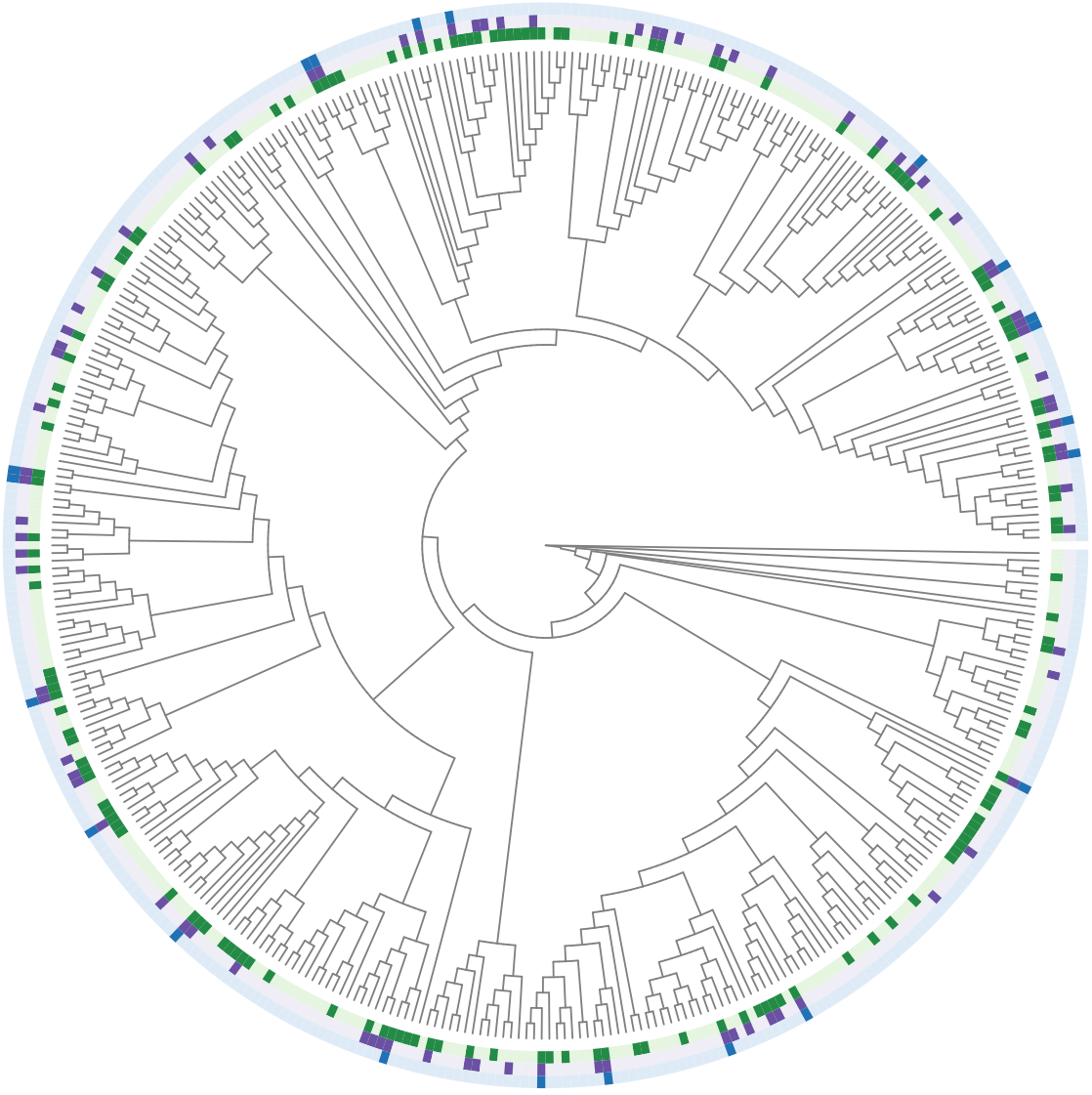
Table 2: Phylogenetic signal measured by mean nearest taxon distance (MNTD) of heavy metal hyperaccumulators and halophytes in phylogenies representing seven angiosperm families. MNTD was evaluated separately for hyperaccumulators and halophytes in each family. Observed MNTD is reported as well as the mean MNTD from 1000 Brownian motion (BM) and 1000 random (ran.) sets. *P*-values indicate whether the observed BMNTD is significantly larger ($p > 0.95$, italics) or significantly smaller ($p < 0.05$, bold) than predicted by the BM or randomized set. Bold text represents values that are significantly smaller than expected (more closely related than expected) under a particular model, and italics show that the observed value is significantly larger than expected (less closely related than expected).

		Heavy metal hyperaccumulators					Halophytes				
Order	Family	MNTD (Obs.)	MNTD (BM)	MNTD (BM *p*)	MNTD (Ran.)	MNTD (Ran. *p*)	MNTD (Obs.)	MNTD (BM)	MNTD (BM *p*)	MNTD (Ran.)	MNTD (Ran. *p*)
Asterales	Asteraceae	10.7	4.8	*0.999*	16.8	**<0.001**	9.5	4.1	*1.000*	12.2	**<0.001**
Brassicales	Brassicaceae	5.0	4.1	0.853	10.2	**<0.001**	7.6	4.9	*0.953*	13.4	**<0.001**
Caryophyllales	Amaranthaceae	14.9	6.1	*0.995*	14.6	0.548	3.2	2.9	*1.000*	3.5	**<0.001**
Fabales	Fabaceae	20.3	7.1	*0.998*	25.2	0.098	8.1	3.8	*1.000*	10.6	**<0.001**
Malpighiales	Euphorbiaceae	13.7	7.2	0.947	18.4	0.067	7.5	5.1	0.909	13.2	**<0.001**
-	Phyllanthaceae	4.2	5.1	0.339	10.5	**<0.001**	11.3	5.9	*0.968*	12.0	0.384
Poales	Poaceae	12.8	5.1	*1.000*	16.1	**0.019**	5.5	3.5	*1.000*	7.5	**<0.001**

Table 3: Results for the between-group mean nearest taxon distances (BMNTD) in phylogenetic trees representing seven angiosperm families. The mean number of estimated species in each family is taken from the APG III website (Stevens 2001). The number of species in each family tree (tips in tree) is stated, along with the number of heavy metal hyperaccumulators (hypers) and halophytes (halos) in each tree, as well as the number of species that are known to be both (referred to as multi-tolerant species, multi). The observed BMNTD is listed as well as the mean BMTD for the 1000 Brownian motion simulations of each trait (BMNTD mean). *P*-values indicate whether the observed BMNTD is smaller ($p < 0.05$) than expected for a model where each trait evolves independently under BM. Bold text represents values that are significantly smaller than expected (more closely related than expected) under BM.

Order	Family	Family size	Tips in tree	Hypers in tree	Halos in tree	Multi in tree	Obs. BMNTD	BMNTD mean	BMNTD (*p*)
Asterales	Asteraceae	23600	4361	40	100	7	16.5	35.9	**0.024**
Brassicales	Brassicaceae	3710	1216	45	21	2	15.2	25.6	0.085
Caryophyllales	Amaranthaceae	2275	580	8	261	6	11.2	20.0	**0.001**
Fabales	Fabaceae	19500	3927	11	133	3	21.9	47.4	**0.014**
Malpighiales	Euphorbiaceae	5735	1030	7	18	1	17.8	27.7	0.116
-	Phyllanthaceae	1745	254	9	6	0	13.2	16.3	0.312
Poales	Poaceae	11160	2101	24	170	12	12.2	29.8	**0.001**

Figure 1: Phylogeny of angiosperm families with halophytes, hyperaccumulators, and multi-tolerant species. The phylogeny contains 401 of the 411 families included in the analysis (see Methods) that are represented in a published phylogeny of angiosperms (Smith et al. 2011). 148 out of the 149 families with halophytes are marked in dark green, all 82 families with heavy metal hyperaccumulators are marked in dark purple, and the 21 families containing multi-tolerant species (able to tolerate salinity and hyperaccumulate heavy metals) are marked in dark blue. The family phylogeny is modified from Saslis-Lagoudakis et al. (2014). Family tip labels are presented in Figure S1. Color labels around the phylogeny were added using the 'trait.plot' function in the R package *diversitree* (FitzJohn 2012).



References

Albert, R. 1975. Salt regulation in halophytes. Oecologia 21:57–71.

Ali, H., E. Khan, & M. A. Sajad. 2013. Phytoremediation of heavy metals—Concepts and applications. Chemosphere 91:869–881.

Anacker, B. L. 2014. The nature of serpentine endemism. American Journal of Botany 101:219–224.

Anderson, C., R. R. Brooks, A. Chiarucci, C. J. LaCoste, M. Leblanc, B. H. Robinson, R. Simcock, & R. B. Stewart. 1999. Phytomining for nickel, thallium and gold. Journal of Geochemical Exploration 67:407–415.

Anjum, N. A., I. Ahmad, M. Válega, I. Mohmood, S. S. Gill, N. Tuteja, A. C. Duarte, & E. Pereira. 2014. Salt marsh halophyte services to metal–metalloid remediation: assessment of the processes and underlying mechanisms. Critical Reviews in Environmental Science and Technology 44:2038–2106.

Antonovics, J., A. D. Bradshaw, & R. G. Turner. 1971. Heavy metal tolerance in plants. Advances in Ecological Research 7:1–85.

APG, III. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society 161:105–121.

Arthur, E. L., P. J. Rice, P. J. Rice, T. A. Anderson, S. M. Baladi, K. L. D. Henderson, & J. R. Coats. 2005. Phytoremediation—an overview. Critical Reviews in Plant Science 24:109–122.

Baker, A. J. M., & R. R. Brooks. 1989. Terrestrial higher plants which hyper-accumulate metallic elements—a review of their distribution, ecology and phytochemistry. Biorecovery 1:81–126.

Baker, A., S. P. McGrath, C. Sidoli, & R. D. Reeves. 1994. The possibility of in situ heavy metal decontamination of polluted soils using crops of metal-accumulating plants. Resources, Conservation and Recycling 11:41–49.

Bartels, D., & R. Sunkar. 2005. Drought and salt tolerance in plants. Critical Reviews in Plant Science 24:23–58.

Bennett, T. H., T. J. Flowers, & L. Bromham. 2013. Repeated evolution of salt-tolerance in grasses. Biology Letters 9:20130029–20130029.

Bert, V., P. Meerts, P. Saumitou-Laprade, P. Salis, W. Gruber, & N. Verbruggen. 2003. Genetic basis of Cd tolerance and hyperaccumulation in *Arabidopsis halleri*. Plant Soil 249:9–18.

Bhatia, N. P., A. J. M. Baker, K. B. Walsh, & D. J. Midmore. 2005. A role for nickel in osmotic adjustment in drought-stressed plants of the nickel hyperaccumulator *Stackhousia tryonii* Bailey. Planta 223:134–139.

Bose, J., A. Rodrigo-Moreno, & S. Shabala. 2014. ROS homeostasis in halophytes in the context of salinity stress tolerance. Journal of Experimental Botany 65:1241–1257.

Boyd, R. S. 2004. Ecology of metal hyperaccumulation. New Phytologist 162:563–567.

Briat, J. F., & M. Lebrun. 1999. Plant responses to metal toxicity. Plant Biology and Pathology 322:43–54.

Bromham, L., & T. H. Bennett. 2014. Salt tolerance evolves more frequently in C_4 grass lineages. Journal of Evolutionary Biology 27:653–659.

Brooks, R. R., M. F. Chambers, L. J. Nicks, & B. H. Robinson. 1998. Phytomining. Trends in Plant Science 3:359–362.

Cappa, J. J., & E. A. H. Pilon-Smits. 2014. Evolutionary aspects of elemental hyperaccumulation. Planta 239:267–275.

Cayuela, L., Í. Granzow-de la Cerda, F. S. Albuquerque, & D. J. Golicher. 2012. Taxonstand: An R package for species names standardisation in vegetation databases. Methods in Ecology and Evolution 3:1078–1083.

Cecchi, L., R. Gabbrielli, M. Arnetoli, C. Gonnelli, A. Hasko, & F. Selvi. 2010. Evolutionary lineages of nickel hyperaccumulation and systematics in European Alysseae (Brassicaceae): evidence from nrDNA sequence data. Annals of Botany 106:751–767.

Colmer, T. D., & T. J. Flowers. 2008. Flooding tolerance in halophytes. New Phytologist 179:964–974.

Colmer, T. D., T. J. Flowers, & R. Munns. 2006. Use of wild relatives to improve salt tolerance in wheat. Journal of Experimental Botany 57:1059–1078.

Dagar, J. C., & S. Gurbachan. 2007. Biodiversity of saline and waterlogged environments: documentation, utilization and management. NBA Scientific Bulletin Number - 9, National Biodiversity Authority, Chennai, Tamil Nadu, India, p78.

de Silva, N. D. G., E. Cholewa, & P. Ryser. 2012. Effects of combined drought and heavy metal stresses on xylem structure and hydraulic conductivity in red maple (*Acer rubrum* L.). Journal of Experimental Botany 63:5957–5966.

Edwards, E. J., & M. J. Donoghue. 2013. Is it easy to move and easy to evolve? Evolutionary accessibility and adaptation. Journal of Experimental Botany 64:4047–4052.

El Shaer, H. M. 2010. Halophytes and salt-tolerant plants as potential forage for ruminants in the Near East region. Small Ruminant Research 91:3–12.

Felsenstein, J. 2005. Using the quantitative genetic threshold model for inferences between and within species. Philosophical Transactions of the Royal Society of London B: Biological Sciences 360:1427–1434.

Feuillet, C., P. Langridge, & R. Waugh. 2008. Cereal breeding takes a walk on the wild side. Trends in Genetics 24:24–32.

FitzJohn, R. G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods in Ecology and Evolution 3:1084–1092.

Flowers, T. J., & A. R. Yeo. 1995. Breeding for salinity resistance in crop plants: where next? Australian Journal of Plant Physiology 22:875–884.

Flowers, T. J., H. K. Galal, & L. Bromham. 2010. Evolution of halophytes: multiple origins of salt tolerance in land plants. Functional Plant Biology 37:604.

Flowers, T., P. F. Troke, & A. R. Yeo. 1977. The mechanism of salt tolerance in halophytes. Annual Review of Plant Physiology 28:89–121.

Fritz, S. A., & A. Purvis. 2010. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. Conservation Biology 24:1042–1051.

Glenn, E. P., J. J. Brown, & E. Blumwald. 1999. Salt tolerance and crop potential of halophytes. Critical Reviews in Plant Science 18:227–255.

Goolsby, E. W., & C. M. Mason. 2015. Toward a more physiologically and evolutionarily relevant definition of metal hyperaccumulation in plants. Frontiers in Plant Science 6:33.

Greenwood, M. E., & G. R. MacFarlane. 2009. Effects of salinity on competitive interactions between two *Juncus* species. Aquatic Botany 90:23–29.

Guvensen, A., G. Gork, & M. Özturk. 2006. An overview of the halophytes in Turkey. In M. A. Khan, G. S. Kust, H.-J. Barth, & B. Böer (Eds.), Sabkha ecosystems. Vol: II. Springer, Netherlands, pp. 9–30.

Hamed, K. B., H. Ellouzi, O. Z. Talbi, K. Hessini, I. Slama, T. Ghnaya, S. M. Bosch, A. Savouré, & C. Abdelly. 2013. Physiological response of halophytes to multiple stresses. Functional Plant Biology 40:883–896.

Hanikenne, M., & C. Nouet. 2011. Metal hyperaccumulation and hypertolerance: a model for plant evolutionary genomics. Current Opinion in Plant Biology 14:252–259.

Haston, E., J. E. Richardson, P. F. Stevens, M. W. Chase, & D. J. Harris. 2009. The Linear Angiosperm Phylogeny Group (LAPG) III: a linear sequence of the families in APG III. Botanical Journal of the Linnean Society 161:128–131.

Hong, Y. 2013. Computational statistics and data analysis. Computational Statistics and Data Analysis 59:41–51.

Hughes, R., K. Bachmann, N. Smirnoff, & M. R. Macnair. 2001. The role of drought tolerance in serpentine tolerance in the *Mimulus guttatus* Fischer ex DC. complex. South African Journal of Science 97:581–586.

Jordan, F., M. Robin Abbott, R. M. Maier, & E. P. Glenn. 2002. A comparison of chelator-facilitated metal uptake by a halophyte and a glycophyte. Environmental Toxicology and Chemistry 21:2698–2704.

Kadukova, J., E. Manousaki, & N. Kalogerakis. 2008. Pb and Cd Accumulation and Phyto-Excretion by Salt Cedar (T*amarix smyrnensis* Bunge). International Journal of Phytoremediation 10:31–46.

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D.

Ackerly, S. P. Blomberg, & C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464.

Khan, M. A., & M. Qaiser. 2006. Halophytes of Pakistan: characteristics, distribution and potential economic usages. In M. A. Khan, G. S. Kust, H.-J. Barth, & B. Böer (Eds.), Sabkha ecosystems. Vol: II. Springer, Netherlands, pp. 129–153.

Krämer, U. 2010. Metal hyperaccumulation in plants. Annual Review of Plant Biology 61:517–534.

Lefèvre, I., G. Marchal, P. Meerts, E. Corréal, & S. Lutts. 2009. Chloride salinity reduces cadmium accumulation by the Mediterranean halophyte species *Atriplex halimus* L. Environmental and Experimental Botany 65:142–152.

Lutts, S., & I. Lefevre. 2015. How can we take advantage of halophyte properties to cope with heavy metal toxicity in salt-affected areas? Annals of Botany 115:509–528.

Mahajan, S., & N. Tuteja. 2005. Cold, salinity and drought stresses: an overview. Archives of Biochemistry and Biophysics 444:139–158.

Manousaki, E., & N. Kalogerakis. 2011a. Halophytes present new opportunities in phytoremediation of heavy metals and saline soils. Industrial and Engineering Chemistry Research 50:656–660.

Manousaki, E., & N. Kalogerakis. 2011b. Halophytes—an emerging trend in phytoremediation. International Journal of Phytoremediation 13:959–969.

Menzel, U., & H. Lieth. 2003. Halophyte database version 2.0. In H. Lieth & M. Mochtchenko (Eds.), Cash crop halophytes: recent studies. Kluwer Academic Publishers, Dordrecht, pp. 221–250.

Moray, C., X. Hua, & L. Bromham. 2015. Salt tolerance is evolutionarily labile in a diverse set of angiosperm families. BMC Evolutionary Biology 15:90.

Munns, R. 2005. Genes and salt tolerance: bringing them together. New Phytologist 167:645–663.

Munns, R., & M. Tester. 2008. Mechanisms of salinity tolerance. Annual Review of Plant Biology. 59:651–681.

Munns, R., R. A. James, & A. Läuchli. 2006. Approaches to increasing the salt tolerance of wheat and other cereals. Journal of Experimental Botany 57:1025–1043.

Nriagu, J. O. 1979. Global inventory of natural and anthropogenic emissions of trace metals to the atmosphere. Nature 279:409–411.

Öztürk, M., A. Guvensen, S. Sakçali, & G. Gork. 2008. Halophyte plant diversity in the Irano-Turanian phytogeographical region of Turkey. In C. Abdelly, M. Öztürk, M. Ashraf, & C. Grignon (Eds.), Biosaline Agriculture and High Salinity Tolerance. Birkhäuser, Basel, pp. 141–155.

Panta, S., T. Flowers, P. Lane, R. Doyle, G. Haros, & S. Shabala. 2014. Halophyte agriculture: success stories. Environmental and Experimental Botany 107:71–83.

Paradis, E., J. Claude, & K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289.

Proctor, J. 1999. Toxins, nutrient shortages and droughts: the serpentine challenge. Trends in Ecology and Evolution 14:334–335.

Przymusiński, R., R. Rucińska, & E. A. Gwóźdź. 2004. Increased accumulation of pathogenesis-related proteins in response of lupine roots to various abiotic stresses. Environmental and Experimental Botany 52:53–61.

Rabhi, M., S. Ferchichi, J. Jouini, M. H. Hamrouni, H.-W. Koyro, A. Ranieri, C. Abdelly, & A. Smaoui. 2010. Phytodesalination of a salt-affected soil with the halophyte *Sesuvium portulacastrum* L. to arrange in advance the requirements for the successful growth of a glycophytic crop. Bioresource Technology 101:6822–6828.

Rascio, N., & F. Navari-Izzo. 2011. Heavy metal hyperaccumulating plants: how and why do they do it? And what makes them so interesting? Plant Science 180:169–181.

Ravindran, K. C., K. Venkatesan, V. Balakrishnan, K. P. Chellappan, & T. Balasubramanian. 2007. Restoration of saline land by halophytes for Indian soils. Soil Biology and Biochemistry 39:2661–2664.

Redondo-Gómez, S. 2013. Bioaccumulation of heavy metals in *Spartina*. Functional Plant Biology 40:913–921.

Redondo-Gómez, S., E. Mateos-Naranjo, & L. Andrades-Moreno. 2010. Accumulation and tolerance characteristics of cadmium in a halophytic Cd-hyperaccumulator, *Arthrocnemum macrostachyum*. Journal of Hazardous Materials 184:299–307.

Reeves, R. D., C. Schwartz, J. L. Morel, & J. Edmondson. 2001. Distribution and metal-accumulating behavior of *Thlaspi caerulescens* and associated metallophytes in France. International Journal of Phytoremediation 3:145–172.

Rozema, J., & T. Flowers. 2008. Crops for a salinized world. Science 322:1478–1480.

Rozema, J., D. Cornelisse, Y. Zhang, H. Li, B. Bruning, D. Katschnig, R. Broekman, B. Ji, & P. van Bodegom. 2014. Comparing salt tolerance of beet cultivars and their halophytic ancestor: consequences of domestication and breeding programs. AoB plants: plu083.

Sage, R. F. 2004. The evolution of C_4 photosynthesis. New Phytologist 161:341–370.

Saslis-Lagoudakis, C. H., C. Moray, & L. Bromham. 2014. Evolution of salt tolerance in angiosperms: a phylogenetic approach. In N. Rajakaruna, R. S. Boyd, and T. B. Harris (Eds.), Plant ecology and evolution in harsh environments. Nova Science Publishers, Hauppauge, NY, pp. 77–95.

Schat, H., S. S. Sharma, & R. Vooijs. 1997. Heavy metal-induced accumulation of free proline in a metal-tolerant and a nontolerant ecotype of *Silene vulgaris*. Physiologia Plantarum 101:477–482.

Sharma, S. S., & K.-J. Dietz. 2006. The significance of amino acids and amino acid-derived molecules in plant responses and adaptation to heavy metal stress. Journal of Experimental Botany 57:711–726.

Sheoran, V., A. S. Sheoran, & P. Poonia. 2009. Phytomining: a review. Minerals Engineering 22:1007–1019.

Smith, S. A., J. M. Beaulieu, A. Stamatakis, & M. J. Donoghue. 2011. Understanding angiosperm diversification using small and large phylogenetic trees. American Journal of Botany 98:404–414.

Stevens, P. F. 2001. Angiosperm Phylogeny Website. 12th edition. http://www.mobot.org/MOBOT/research/ APweb/.

Stewart, G. R., & J. A. Lee. 1974. The role of proline accumulation in halophytes. Planta 120:279–289.

Tester, M., & P. Langridge. 2010. Breeding technologies to increase crop production in a changing world. Science 327:818.

The Plant List. 2010. The Plant List. Version 2. http://theplantlist.org.

Thomas, J. C., F. K. Malick, C. Endreszl, E. C. Davies, & K. S. Murray. 1998. Distinct responses to copper stress in the halophyte *Mesembryanthemum crystallinum*. Physiologia Plantarum 102:360–368.

Vara Prasad, M. N., & H. M. de Oliveira Freitas. 2003. Metal hyperaccumulation in plants: biodiversity prospecting for phytoremediation technology. Electronic Journal of Biotechnology 6:285–321.

Ventura, Y., A. Eshel, D. Pasternak, & M. Sagi. 2015. The development of halophyte-based agriculture: past and present. Annals of Botany 115:529–540.

Ventura, Y., & M. Sagi. 2013. Halophyte crop cultivation: the case for *Salicornia* and *Sarcocornia*. Environmental and Experimental Botany 92:144–153.

Webb, C. O., D. D. Ackerly, M. A. McPeek, & M. J. Donoghue. 2002. Phylogenies and community ecology. Annual Review of Ecology and Systematics 33:475–505.

Weber, D. J., R. Ansari, B. Gul, & M. Ajmal Khan. 2007. Potential of halophytes as source of edible oil. Journal of Arid Environments 68:315–321.

Wu, L., A. D. Bradshaw, & D. A. Thurman. 1975. The potential for evolution of heavy metal tolerance in plants. III. The rapid evolution of copper tolerance in *Agrostis stolonifera*. Heredity 34:165–187.

Wuana, R. A., & F. E. Okieimen. 2011. Heavy metals in contaminated soils: a review of sources, chemistry, risks and best available strategies for remediation. ISRN Ecology 2011:1–20.

Zhao, K., J. Song, G. Feng, M. Zhao, & J. Liu. 2011. Species, types, distribution, and economic potential of halophytes in China. Plant Soil 342:495–509.

Supplemental material for "**The phylogenetic association between salt tolerance and heavy metal hyperaccumulation in angiosperms"**

Camile Moray[1], Eric W. Goolsby[2], Lindell Bromham[1]

[1] Macroevolution and Macroecology Group, Division of Evolution, Ecology and Genetics, Research School of Biology, Australian National University, Canberra, Australia

[2]Odum School of Ecology Interdisciplinary Toxicology Program, University of Georgia, Athens, USA

Contents:

Extraction of non-monophyletic family subtrees

We generated a phylogeny for each of the target families (see Methods) by extracting all tips corresponding to each family from a published angiosperm phylogeny (Smith et al. 2011). However, for some families, not all tips were monophyletic in the published angiosperm phylogeny (Smith et al. 2011). For these families we extracted all tips associated with the family, excluding a small number of tips that were either nested within clades of other families or tips from other families that were nested within the clade of the target family. Here we list details on each non-monophyletic family subtree, referring to specific tip numbers associated with the original published phylogeny. For the Asteraceae subtree, we excluded one Asteraceae tip that was nested within the Campanulaceae clade (tip number 10079). For Brassicaceae we excluded five tips that were in the Capparaceae clade (42135, 42136, 42145, 42168, 42171). For Fabaceae, we excluded one Fabaceae tip that was in the Connaraceae clade (34282) removed one Apocynaceae tip (30884). For Euphorbiaceae we excluded four Peraceae tips that were nested in the Euphorbiaceae clade (36371:36374). For Phyllanthaceae we removed two tips from Picrodendraceae (36318, 36319) and one from Putranjivaceae (36312).

Table S1: List of angiosperm families with halophytes and hyperaccumulators. Family size represents the mean estimated number of species in each family according to the APG III website (Stevens 2001). The multi-tolerant spp column lists the number of species in each family that appear on both the list of halophytes and the list of hyperaccumulators.

Order	Family	Family size	Halophytes	Hyperaccumulators	Multi-tolerant spp
Alismatales	Alismataceae	88	3	0	0
-	Aponogetonaceae	43	2	0	0
-	Araceae	4759	8	4	2
-	Butomaceae	1	1	0	0
-	Cymodoceaceae	16	15	0	0
-	Hydrocharitaceae	116	22	0	0
-	Juncaginaceae	15	7	0	0
-	Posidoniaceae	9	3	0	0
-	Potamogetonaceae	102	6	1	0
-	Ruppiaceae	6	2	0	0
-	Zosteraceae	14	17	0	0
Apiales	Apiaceae	3780	34	1	0
-	Araliaceae	1450	3	2	0
Arecales	Arecaceae	2361	37	0	0
Asparagales	Amaryllidaceae	1605	15	1	0
-	Asparagaceae	2480	24	1	0
-	Iridaceae	2025	10	5	2
-	Orchidaceae	22075	3	0	0
-	Xanthorrhoeaceae	900	3	0	0
Asterales	Argophyllaceae	21	0	2	0
-	Asteraceae	23600	275	84	9
-	Calyceraceae	60	1	0	0
-	Campanulaceae	2380	0	2	0
-	Goodeniaceae	430	6	0	0
-	Menyanthaceae	58	2	0	0
Brassicales	Bataceae	2	2	0	0
-	Brassicaceae	3710	40	92	3
-	Capparaceae	480	10	0	0
-	Cleomaceae	300	7	0	0
-	Resedaceae	75	5	1	0
-	Salvadoraceae	11	4	0	0
Buxales	Buxaceae	70	0	15	0
Caryophyllales	Aizoaceae	2035	46	2	2
-	Amaranthaceae	2275	508	11	7
-	Anacampserotaceae	32	1	0	0
-	Basellaceae	19	2	0	0
-	Cactaceae	1866	11	0	0
-	Caryophyllaceae	2200	27	10	0
-	Didiereaceae	16	2	0	0
-	Frankeniaceae	90	17	0	0
-	Gisekiaceae	5	1	0	0
-	Halophytaceae	1	1	0	0
-	Lophiocarpaceae	6	1	0	0
-	Molluginaceae	87	4	1	0
-	Nyctaginaceae	395	9	1	0
-	Phytolaccaceae	65	0	2	0
-	Plumbaginaceae	836	62	1	1
-	Polygonaceae	1110	41	7	1
-	Portulacaceae	70	12	0	0
-	Sarcobataceae	2	1	0	0
-	Simmondsiaceae	1	1	0	0
-	Stegnospermataceae	3	2	0	0

Order	Family	Family size	Halophytes	Hyperaccumulators	Multi-tolerant spp
Caryophyllales	Talinaceae	27	1	1	0
-	Tamaricaceae	90	55	0	0
Celastrales	Celastraceae	1400	9	1	0
Ceratophyllales	Ceratophyllaceae	6	1	0	0
Commelinales	Commelinaceae	652	4	3	0
-	Philydraceae	5	0	1	0
-	Pontederiaceae	33	3	1	1
Cornales	Loasaceae	265	1	0	0
Cucurbitales	Cucurbitaceae	960	14	0	0
Dilleniales	Dilleniaceae	355	1	0	0
Dipsacales	Caprifoliaceae	890	2	1	0
Ericales	Balsaminaceae	1001	0	1	0
-	Ebenaceae	548	4	0	0
-	Ericaceae	3995	1	1	0
-	Lecythidaceae	310	4	1	0
-	Primulaceae	2590	14	1	0
-	Sapotaceae	1100	2	2	0
-	Symplocaceae	320	0	2	0
-	Tetrameristaceae	5	1	0	0
-	Theaceae	328	0	1	0
Fabales	Fabaceae	19500	252	27	4
-	Polygalaceae	965	3	1	0
-	Surianaceae	8	1	0	0
Fagales	Betulaceae	145	1	1	0
-	Casuarinaceae	95	12	0	0
Gentianales	Apocynaceae	4555	44	2	1
-	Gentianaceae	1655	13	0	0
-	Loganiaceae	420	1	0	0
-	Rubiaceae	13150	15	12	0
-	Geraniaceae	805	1	0	0
Lamiales	Acanthaceae	4000	19	2	0
-	Bignoniaceae	800	9	0	0
-	Lamiaceae	7173	31	12	2
-	Linderniaceae	195	2	2	0
-	Oleaceae	615	0	1	0
-	Orobanchaceae	2060	17	7	0
-	Pedaliaceae	70	1	0	0
-	Phrymaceae	188	4	0	0
-	Plantaginaceae	1900	35	2	1
-	Scrophulariaceae	1800	14	1	0
-	Verbenaceae	918	15	3	0
Laurales	Hernandiaceae	55	2	0	0
-	Lauraceae	2500	2	0	0
Liliales	Colchicaceae	245	2	0	0
-	Liliaceae	610	2	0	0
Magnoliales	Annonaceae	2220	1	0	0
-	Myristicaceae	475	0	1	0
Malpighiales	Bonnetiaceae	35	1	0	0
-	Chrysobalanaceae	460	2	0	0
-	Clusiaceae	595	2	5	0
-	Dichapetalaceae	165	0	1	0
-	Elatinaceae	35	7	0	0
-	Euphorbiaceae	5735	43	37	3
-	Hypericaceae	560	1	0	0
-	Linaceae	300	4	0	0
-	Ochnaceae	495	0	3	0
-	Passifloraceae	935	0	5	0
-	Phyllanthaceae	1745	9	39	0
-	Putranjivaceae	210	1	0	0
-	Rhizophoraceae	149	19	0	0
Malpighiales	Salicaceae	1010	6	2	0
Malpighiales	Violaceae	800	1	4	0

Order	Family	Family size	Halophytes	Hyperaccumulators	Multi-tolerant spp
Malvales	Cistaceae	175	1	1	0
-	Malvaceae	4225	56	8	2
-	Neuradaceae	10	1	0	0
-	Thymelaeaceae	891	3	1	0
Myrtales	Combretaceae	500	12	0	0
-	Lythraceae	620	23	2	1
-	Melastomataceae	5005	1	3	0
-	Myrtaceae	4620	51	12	0
-	Onagraceae	656	6	0	0
-	Vochysiaceae	190	0	4	0
Nymphaeales	Nymphaeaceae	58	3	0	0
Oxalidales	Cunoniaceae	280	0	1	0
-	Oxalidaceae	770	2	1	0
Pandanales	Pandanaceae	885	11	0	0
-	Velloziaceae	240	0	1	0
Picramniales	Picramniaceae	49	1	0	0
Piperales	Piperaceae	3615	1	0	0
-	Saururaceae	6	1	1	0
Poales	Bromeliaceae	1770	2	0	0
-	Cyperaceae	5430	124	8	1
-	Flagellariaceae	4	1	0	0
-	Juncaceae	430	22	1	0
-	Poaceae	11160	345	29	14
-	Restionaceae	500	2	0	0
-	Typhaceae	25	11	0	0
Proteales	Nelumbonaceae	2	1	0	0
-	Proteaceae	1600	0	2	0
Ranunculales	Menispermaceae	442	3	0	0
-	Papaveraceae	760	3	2	0
-	Ranunculaceae	2525	18	1	0
Rosales	Elaeagnaceae	45	3	0	0
-	Moraceae	1125	7	0	0
-	Rhamnaceae	925	6	0	0
-	Rosaceae	2520	10	1	0
-	Ulmaceae	35	1	0	0
-	Urticaceae	2625	0	1	0
Santalales	Olacaceae	57	1	0	0
-	Santalaceae	990	3	0	0
Sapindales	Anacardiaceae	873	7	0	0
-	Meliaceae	615	6	0	0
-	Nitrariaceae	16	8	0	0
-	Rutaceae	2070	5	0	0
-	Sapindaceae	1630	3	1	0
-	Simaroubaceae	110	1	0	0
Saxifragales	Crassulaceae	1370	2	6	0
-	Cynomoriaceae	2	1	0	0
-	Haloragaceae	145	0	1	0
Solanales	Convolvulaceae	1625	22	5	1
-	Hydroleaceae	12	1	0	0
-	Solanaceae	2460	42	2	1
Unplaced	Boraginaceae	2755	37	2	0
Vitales	Vitaceae	850	4	0	0
Zingiberales	Zingiberaceae	1208	1	0	0
Zygophyllales	Zygophyllaceae	285	30	1	1

Table S2: Angiosperm species identified as both halophytes and hyperaccumulators. Species names follow The Plant List (2010) taxonomy, and order and family taxonomy follows APG III (2009).

Order	Family	Species
Alismatales	Araceae	*Pistia stratiotes*
-	-	*Spirodela polyrrhiza*
Asparagales	Iridaceae	*Iris ensata*
-	-	*Iris lactea*
Asterales	Asteraceae	*Ageratum conyzoides*
-	-	*Centaurea virgata*
-	-	*Cirsium arvense*
-	-	*Dittrichia viscosa*
-	-	*Erigeron canadensis*
-	-	*Lactuca orientalis*
-	-	*Sonchus arvensis*
-	-	*Sonchus asper*
-	-	*Taraxacum mongolicum*
Brassicales	Brassicaceae	*Alyssum pateri*
-	-	*Brassica juncea*
-	-	*Raphanus raphanistrum*
Caryophyllales	Aizoaceae	*Mesembryanthemum crystallinum*
-	-	*Sesuvium portulacastrum*
-	Amaranthaceae	*Amaranthus hybridus*
-	-	*Arthrocnemum macrostachyum*
-	-	*Atriplex confertifolia*
-	-	*Atriplex halimus*
-	-	*Beta vulgaris*
-	-	*Salsola kali*
-	-	*Salsola soda*
-	Plumbaginaceae	*Armeria maritima*
-	Polygonaceae	*Polygonum aviculare*
Commelinales	Pontederiaceae	*Eichhornia crassipes*
Fabales	Fabaceae	*Aeschynomene indica*
-	-	*Melilotus officinalis*
-	-	*Prosopis laevigata*
-	-	*Tephrosia villosa*
Gentianales	Apocynaceae	*Hemidesmus indicus*
Lamiales	Lamiaceae	*Clerodendrum infortunatum*
-	-	*Ocimum tenuiflorum*
-	Plantaginaceae	*Plantago asiatica*
Malpighiales	Euphorbiaceae	*Croton bonplandianus*
-	-	*Euphorbia hirta*
-	-	*Euphorbia macroclada*
Malvales	Malvaceae	*Alcea rosea*
-	-	*Sida rhombifolia*
Myrtales	Lythraceae	*Trapa natans*
Poales	Cyperaceae	*Schoenoplectus americanus*
-	Poaceae	*Agrostis stolonifera*
-	-	*Chrysopogon zizanioides*
-	-	*Cynodon dactylon*
-	-	*Dactyloctenium aegyptium*
-	-	*Desmostachya bipinnata*
-	-	*Dichanthium annulatum*
-	-	*Echinochloa stagnina*
-	-	*Eleusine indica*
-	-	*Festuca arundinacea*
-	-	*Festuca ovina*
-	-	*Imperata cylindrica*
-	-	*Melinis repens*
-	-	*Paspalum conjugatum*
-	-	*Spartina spartinae*

Order	Family	Species
Solanales	Convolvulaceae	*Evolvulus alsinoides*
-	Solanaceae	*Solanum americanum*
Zygophyllales	Zygophyllaceae	*Zygophyllum fabago*

Figure S1: Phylogeny of angiosperm families containing 401 of the 411 families included in the analysis (see Methods) that are represented in a published phylogeny of angiosperms (Smith et al. 2011). 148 out of the 149 families with halophytes are marked in dark green, all 82 families with heavy metal hyperaccumulators are marked in dark purple, and the 21 families containing multi-tolerator species (able to tolerate salinity and hyperaccumulate heavy metals) are marked in dark blue. The family phylogeny is modified from Saslis-Lagoudakis et al. (2014).

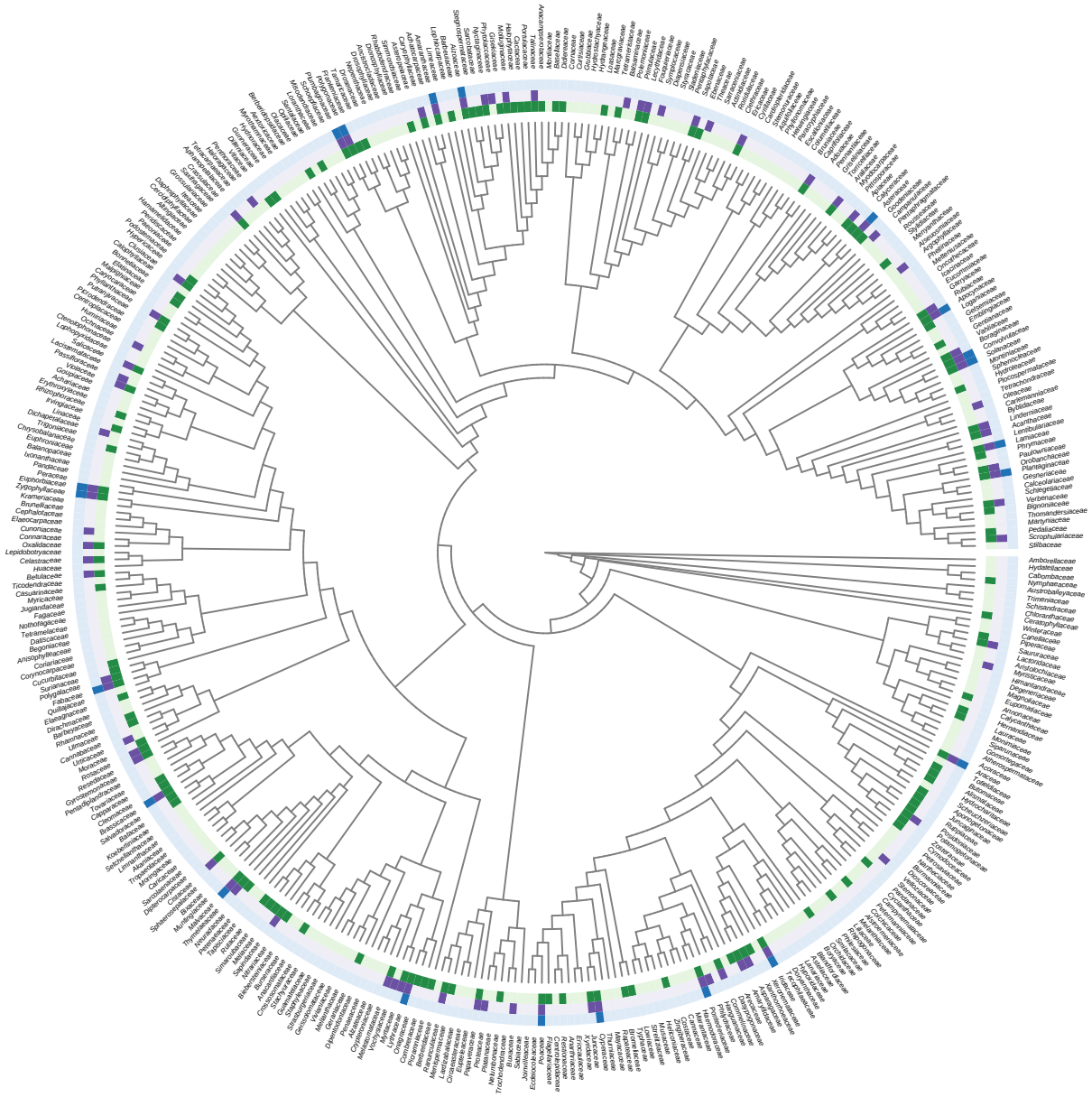
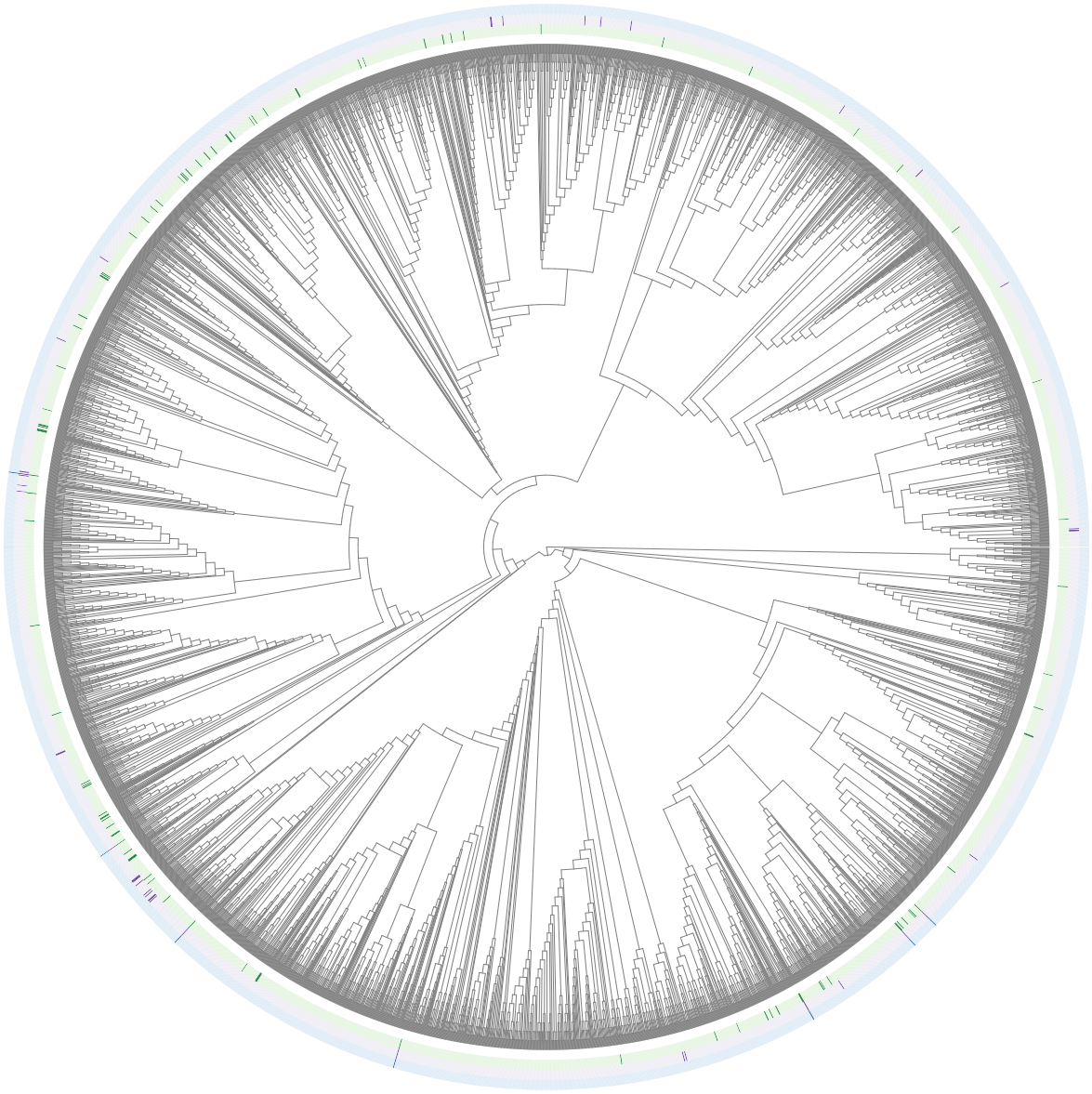
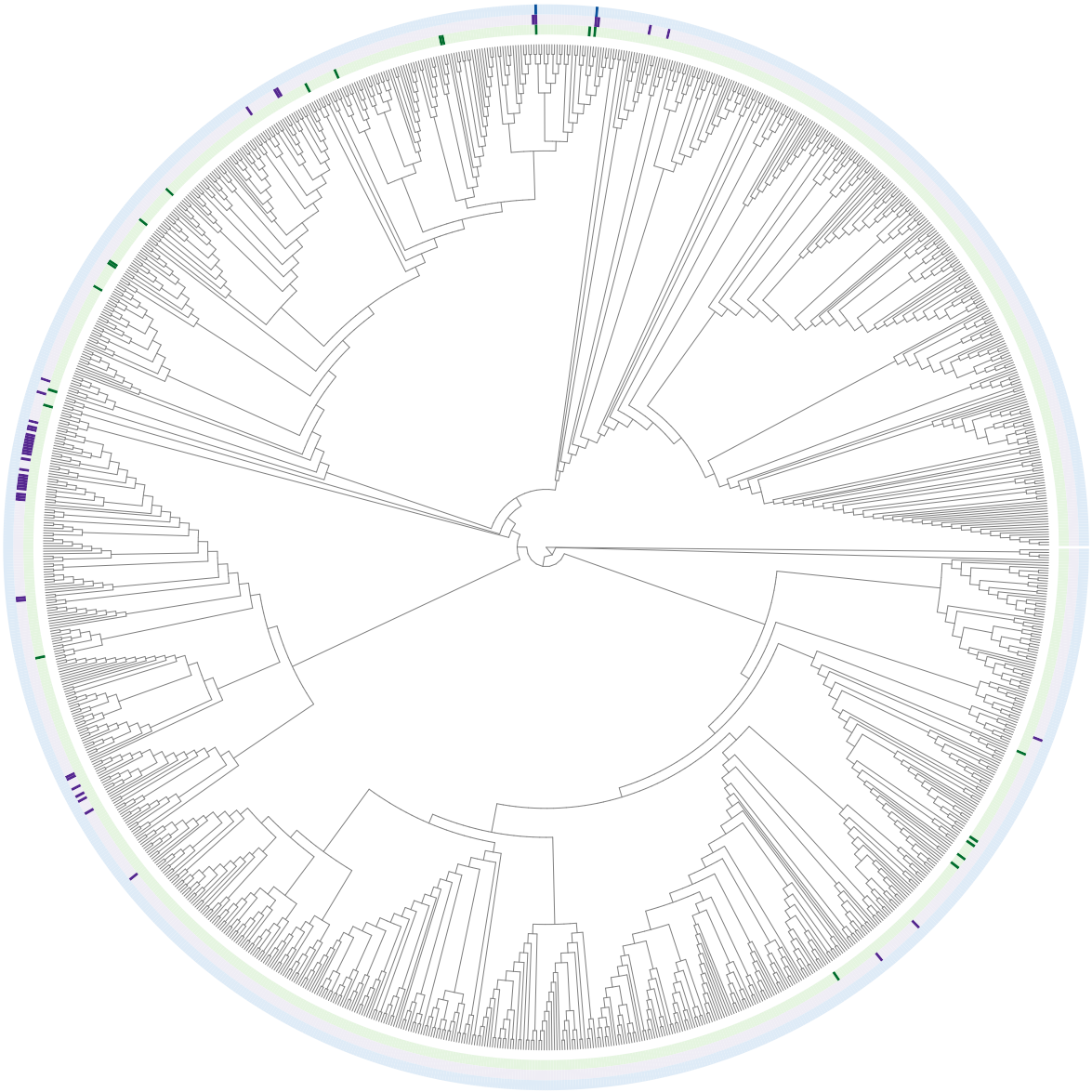


Figure S2: Phylogenies for the sample of seven angiosperm families analyzed (a-g). Tips in the phylogenies identified as halophytes are marked in dark green, tips identified as hyperaccumulators are marked in dark purple, and tips identified as both a halophyte and hyperaccumulator are marked in dark blue in the ring around each phylogeny. Color labels around each phylogenies were added using the 'trait.plot' function in the R package *diversitree* (FitzJohn 2012).

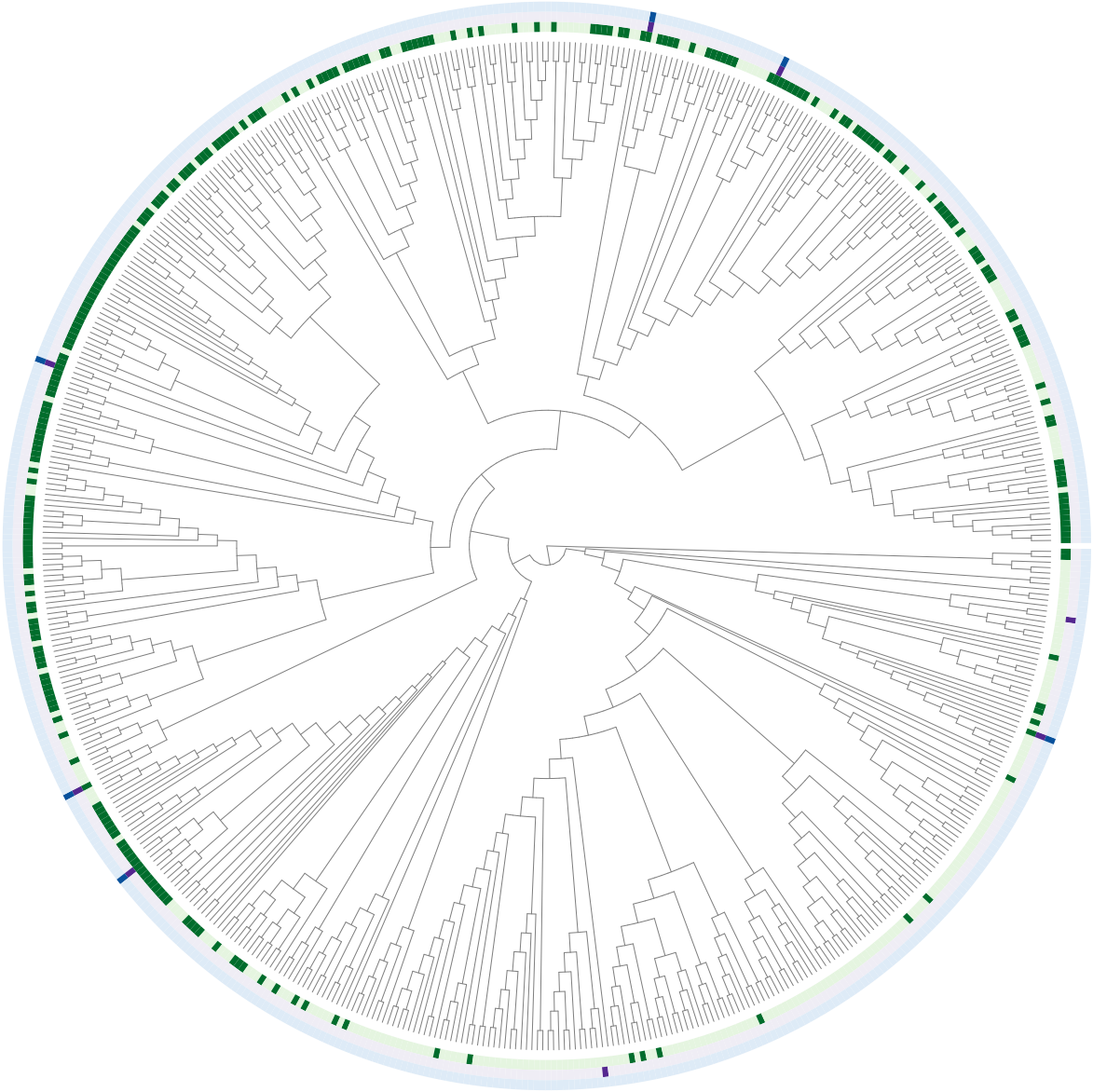
a) Asteraceae



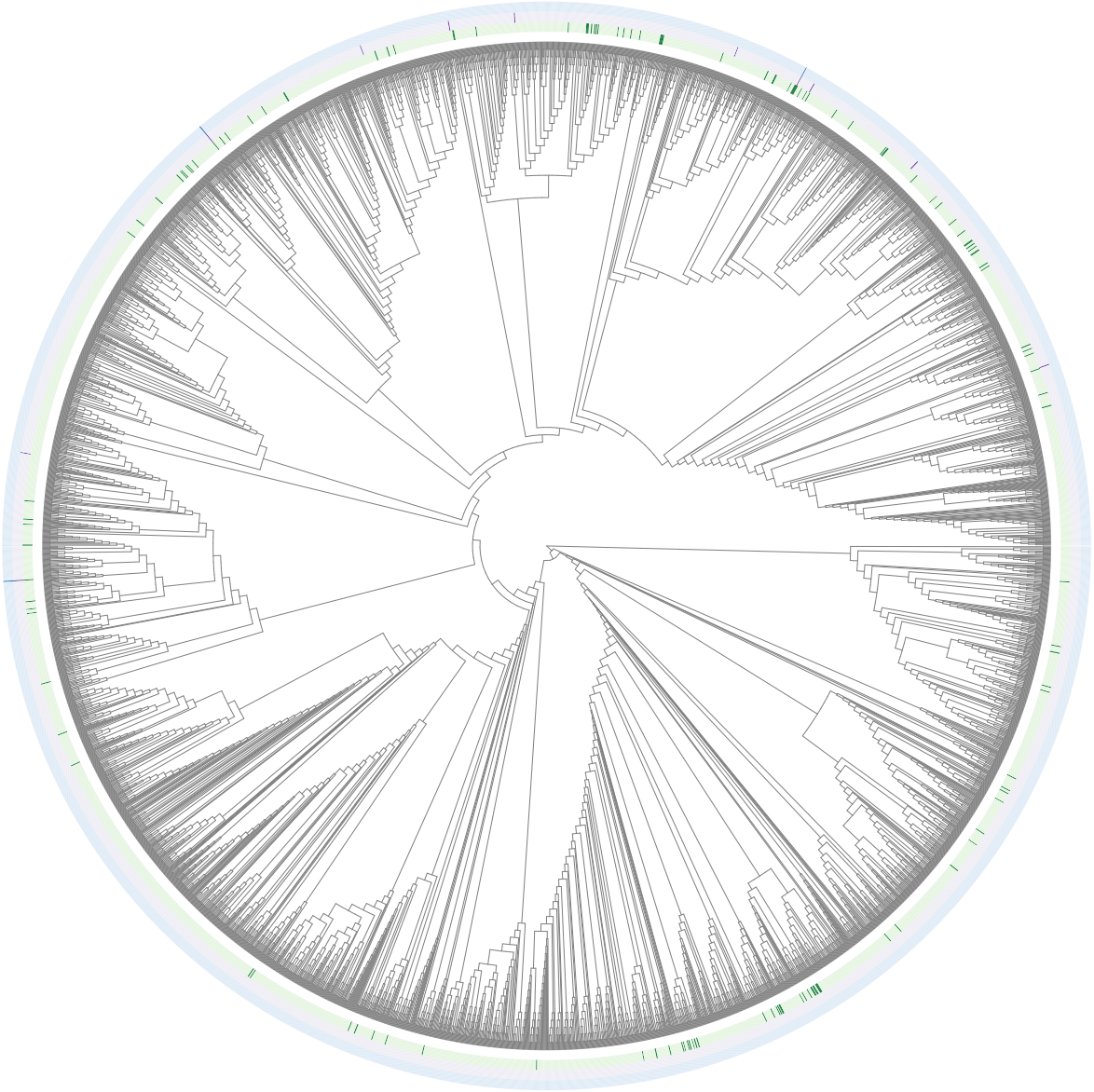
b) Brassicaceae



c) Amaranthaceae



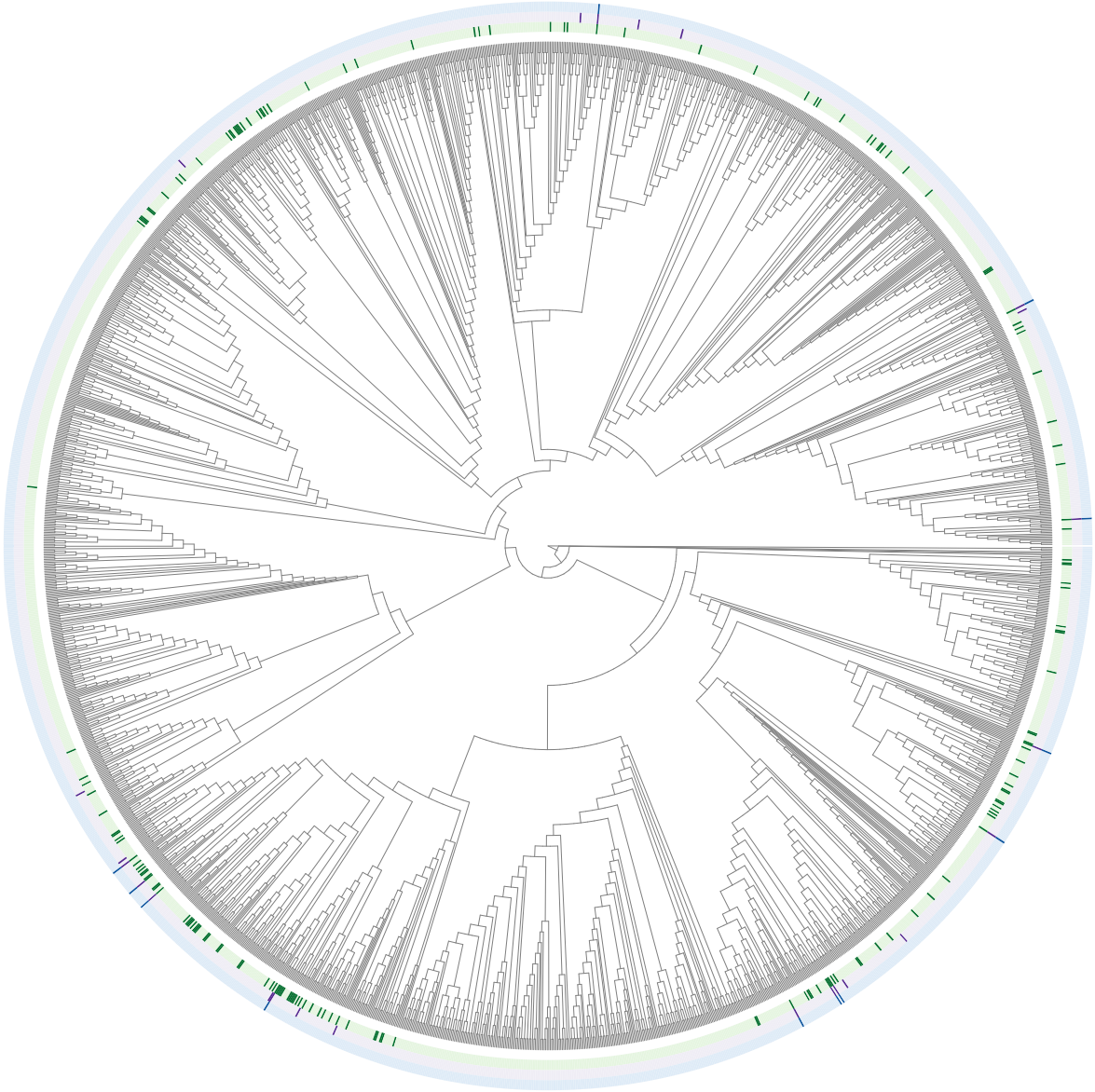
d) Fabaceae



e) Euphorbiaceae

f) Phyllanthaceae

g) Poaceae



References

APG, III. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society 161:105–121.

FitzJohn, R. G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods in Ecology and Evolution 3:1084–1092.

Saslis-Lagoudakis, C. H., C. Moray, & L. Bromham. 2014. Evolution of salt tolerance in angiosperms: a phylogenetic approach. In N. Rajakaruna, R. S. Boyd, and T. B. Harris (Eds.), Plant ecology and evolution in harsh environments. Nova Science Publishers, Hauppauge, NY, pp. 77–95.

Smith, S. A., J. M. Beaulieu, A. Stamatakis, & M. J. Donoghue. 2011. Understanding angiosperm diversification using small and large phylogenetic trees. American Journal of Botany 98:404–414.

Stevens, P. F. 2001. Angiosperm Phylogeny Website. 12th edition. http://www.mobot.org/MOBOT/research/ APweb/.

The Plant List. 2010. The Plant List. Version 2. http://theplantlist.org.

Chapter 4

Evidence of increased rates of molecular evolution in mitochondrial genes of salt tolerant plants

Camile Moray and Lindell Bromham

Abstract

A rare number of organisms from all domains of life have evolved the ability to tolerate and even thrive in saline habitats, which are toxic to many organisms. But exposure to saline habitats incurs costs, both indirectly due to investment in tolerance mechanisms as well as the direct cost of oxidative damage triggered from exposure to salinity. Oxidative damage can lead to DNA damage, which may increase the rate of molecular evolution. Examples in Archaea, bacteria, and crustaceans suggest that saline-adapted lineages have increased rates of molecular evolution, which could be due to the mutagenic effect of salinity, though not all studies have found this pattern. Here, we test whether there is a consistent effect of salinity on the rates of molecular evolution in flowering plants by comparing 17 salt tolerant plants and their non-salt tolerant relatives. We find evidence that salt tolerant plants have faster total substitution rates in mitochondrial genes, but not in chloroplast or nuclear genes. One explanation for this result is that increased total substitution rates are a general feature of salt tolerant lineages, consistent with an effect of oxidative damage associated with exposure to salinity. This study adds to the growing literature on the effects of environmental factors on rates of molecular evolution.

Introduction

Several studies have reported that organisms living in saline habitats have higher rates of molecular evolution compared to closely related non-salt tolerant lineages, including crustaceans (Hebert et al. 2002; Wägele et al. 2003; Baxevanis et al. 2006), Archaea (Dennis and Shimmin 1997), and bacteria (Logares et al. 2010). This genetic signature suggests that some aspects of salt tolerance or living in a saline habitat may increase rates of molecular evolution, the pace at which DNA sequences accumulate changes. However, not all studies on molecular rates in organisms living in saline habitats have found an increase compared to relatives in non-saline habitats. One study on prokaryotes (Logares et al. 2009) and another on plants (Whittle 2006) found no difference in molecular rates between salt tolerant and non-salt tolerant relatives. Understanding whether living in saline habitats can consistently influence molecular evolution is important for establishing the impact of environmental factors on the rate of genome evolution.

We focus on salt tolerance in angiosperms as a study system to analyze the association between salt tolerance and rates of molecular evolution. Salt tolerant plants (halophytes) are relatively rare, representing only 1-2% of plant species (Glenn et al. 1999; Flowers and Colmer 2008). Although they are rare, halophytes are found in a diverse range of angiosperm lineages, and salt tolerance has evolved a very large number of times in the angiosperms (Flowers et al. 2010; Bennett et al. 2013; Moray et al. submitted). The presence of multiple evolutionary origins of salt tolerance gives us the opportunity to examine the effect of salinity or salt tolerance on rates of molecular evolution in a comparative phylogenetic framework, by comparing salt tolerant lineages to their non-salt tolerant relatives. Mechanisms and levels of salinity tolerance vary among halophytes (Flowers et al. 1977). This means that by examining rates of molecular evolution in a variety of halophytes from diverse lineages, we can test the general effect of salinity or salt tolerance, rather than the influence of a particular mechanism or habitat with a particular level of salinity.

Salinity triggers oxidative stress through the production of reactive oxygen species (ROS) in many organisms (Martínez-Alvarez et al. 2002; Parida and Das 2005; Lushchak 2011). In plants, ROS (e.g., hydrogen peroxide and superoxide anion) are

regularly produced in chloroplasts, mitochondria, and peroxisomes as a product of photosynthesis and aerobic respiration. If uncontrolled, ROS can damage lipids, proteins, and nucleic acids (Apel and Hirt 2004; Navrot et al. 2007; Roldán-Arjona and Ariza 2009). To prevent this damage, plants have pathways that limit ROS production or neutralize them (e.g., antioxidants, Mittler 2002; Halliwell 2006; Gill and Tuteja 2010). However, ROS production can increase substantially due to abiotic or biotic stresses, exceeding a plant's capacity for mitigating the effects of ROS (Rhoads 2006). If ROS production exceeds normal levels, for example under salinity stress (Ozgur et al. 2013; Bose et al. 2014), plants may be unable to prevent damage, potentially resulting in damage to DNA. Research suggests that halophytes experience increased ROS exposure and damage (Lechno et al. 1997; Møller 2001; Song et al. 2006; Gong et al. 2010). If prolonged exposure to salinity in halophytes leads to repeated ROS damage (Mittler et al. 2004), halophytes might experience more mutations, particularly in the chloroplast and mitochondrial genomes, where ROS are most often produced. In support of this, one recent study suggests that exposure to salinity in the model species *Arabidopsis thaliania* increases the accumulation of mutations (Jiang et al. 2014).

An increase in the mutation rate as a consequence of ROS damage may influence rates of molecular evolution in halophytes, both directly and indirectly. A direct effect can occur if DNA damage from ROS causes nucleotide changes that are unrepaired or repaired imperfectly, as those changes can be passed on to copies made of the genome. Indirectly, ROS might cause nucleotide changes in genes for DNA control and repair mechanisms, possibly reducing the fidelity of future DNA replication and repair (Lynch 2010) and leading to further mutations. If mutations occur in cells that produce gametes, then they can be inherited. Since plants do not have dedicated germ lines, mutations in somatic tissues can also be inherited (Klekowski 1988; Klekowski and Godfrey 1989).

In the absence of selection, an increase in the mutation rate can increase the substitution rate, the rate at which changes in nucleotide sequences become predominant in a population, as the generation of more mutations increases the chance that some will become ubiquitous in the population. While most deleterious mutations will be removed by selection, synonymous changes in protein coding genes (changes that do not change the amino acid sequence) should experience relatively low selective pressure. This means that the mutation rate should be approximately proportional to the rate of synonymous changes (d_s) (Kimura 1983). If ROS damage in halophytes leads to

increased mutation rates, we are most likely to detect this signal as increased d_s. This effect might be more pronounced in the mitochondrial and chloroplast genomes where ROS are likely to be produced during photosynthesis and respiration. An increased mutation rate increases the rate at which variants are generated, so can also contribute to the substitution rate of non-synonymous sites (d_N), which lead to changes in amino acids. However, d_N will also reflect changes in selection pressures and population size (Kimura 1983), which may be associated with transitions to saline habitats. Increased mutation rates can also lead to changes in the total substitution rate, which includes substitutions in non-coding genes and both synonymous and non-synonymous substitutions in protein-coding genes.

However, many halophytes have adaptations to remove or control ROS (Lechno et al. 1997; Parida and Jha 2009; Bose et al. 2014; Uzilday et al. 2015), which may mitigate effects of increased oxidative damage. Many halophytes have enhanced ROS scavenging and antioxidant activity (Cai-Hong et al. 2005; Seckin et al. 2010; Ozgur et al. 2013; Bose et al. 2014). They may also produce osmoprotectants, like proline, which can protect cellular structures from oxidative damage and can reduce or mitigate ROS production (Stewart and Lee 1974; Munns and Tester 2008; Slama et al. 2015). The enhanced protection against ROS in halophytes might explain why a previous study did not find a difference in molecular rates between halophytes and their relatives in the nuclear 18S ITS region (Whittle 2006). Furthermore, increasing evidence suggests that ROS do not solely contribute to cell damage, but also play a paramount role in stress signaling pathways (Apel and Hirt 2004; Foyer and Noctor 2005; Fujita et al. 2006; Miller et al. 2008; Mittler et al. 2011). When ROS are produced from exposure to excess salinity, they can function as a substrate to activate pathways for ROS scavenging and antioxidant systems (Ozgur et al. 2013; Bose et al. 2014), and halophytes may be more effective at both utilizing ROS for stress signaling and neutralizing them once signaling is complete (Ellouzi et al. 2011; Bose et al. 2014). It has even been suggested that these mechanisms are so efficient in some halophytes that salinity exposure does not induce the production of toxic amounts of ROS (Ellouzi et al. 2011; Bose et al. 2014). Thus, plants must reach a balance between allowing ROS production to trigger protection systems and preventing a level of ROS accumulation that exceeds their mitigation capabilities (Rhoads 2006). And the added stress of salinity might make finding this balance more precarious.

Here we aim to test whether living in saline habitats consistently increases rates of molecular evolution, using halophytes as a study system. We use a comparative phylogenetic approach to test whether salt tolerance influences rates of molecular evolution in angiosperms. We compare rates of molecular evolution between 17 halophytes and related non-halophytes from a diverse range of angiosperm lineages. Specifically, we compare the rate of non-synonymous substitutions that lead to amino acid changes (d_N), the rate of synonymous substitutions that do not change amino acids (d_S), the ratio of non-synonymous to synonymous changes (d_N/d_S), and the total substitution rate in sequences from the mitochondrial, chloroplast and nuclear genomes in order to assess possible causes of changes in rates of molecular evolution.

Methods

Data collection

To test whether halophytes have faster rates of molecular evolution than their non-halophyte relatives, we collected a set of phylogenetically independent comparisons, each including one halophyte, one non-halophyte relative, and one outgroup taxon. For each triplet comparison we collected gene sequences that were available for both the ingroup and outgroup taxa, so that we could compare estimates of rates of molecular evolution between each halophyte and non-halophyte relative, in comparison to the outgroup.

To identify comparisons, we first identified halophytes using the eHALOPH electronic database of halophytes (www.sussex.ac.uk/affiliates/halophytes/, v.3.06). This database contains a list of halophytes, defined as species that can complete their life cycle in a saline habitat. To ensure that we collected independent comparisons, we first searched for the accepted taxonomic names of the halophytes on The Plant List (2010). Mitochondrial genes are available for fewer species than chloroplast and nuclear genes, and because they evolve more slowly than chloroplast or nuclear sequences in plants (Wolfe et al. 1987; Knoop 2004; Galtier 2011), they often do not vary as much between close relatives. The identification of appropriate mitochondrial comparisons was thus the limiting factor in data collection. To account for these restrictions, we prioritized

collection of halophytes with the most mitochondrial genes available on GenBank (www.ncbi.nlm.nih.gov/genbank/). First, we identified halophyte species with sequences available on GenBank. We then identified which of these species had mitochondrial sequences available by searching GenBank with the function 'ncbi_search' in the R package *taxize* (Chamberlain and Socz 2013; Chamberlain *et al*., 2014). For genera with multiple halophytes, we chose to focus on the halophyte with the highest number of mitochondrial genes available on GenBank.

For each identified halophyte with available mitochondrial genes, we then searched for appropriate sequences from a closely related non-halophyte species. Species were considered non-halophytes if they were not included in the eHALOPH electronic database of halophytes and a literature search recovered no evidence of salt tolerance. Prioritizing the collection of mitochondrial data led to the collection of fewer comparisons and some more distantly related comparisons than if we had only collected data from other genomes. In addition to the limited data available for mitochondrial genes on GenBank, we were also limited by requiring comparisons with a sufficient number of changes between the aligned halophyte and non-halophyte sequences to estimate differences in molecular rates. For some comparisons, mitochondrial sequences were available for closely related non-halophytes, but had no or very few synonymous and non-synonymous changes. Since rate estimations are based on counts of substitutions in the sequences in a comparison, in some cases we needed to choose more distantly related non-halophytes in order to have a sufficient number of changes to detect differences in rates between the halophyte and non-halophyte.

For each comparison of a halophyte and non-halophyte relative, we also chose one outgroup species with matching sequence data. In this framework, we can assume that each halophyte and non-halophyte comparison has had equal time to accumulate changes in DNA sequences since their last common ancestor, and those changes can then be compared to each other and to the outgroup species to estimate and compare rates of molecular evolution (Lanfear et al. 2010). Where possible, we also collected chloroplast and nuclear sequences for each comparison. A complete list of genes collected is available in the Supplementary Material, and a summary of alignment lengths is presented in Table 1.

Alignments

For each triplet comparison, we aligned gene sequences across the halophyte, non-halophyte and outgroup species. We constructed separate alignments for each comparison for the sequences from the mitochondrial, chloroplast, and nuclear genomes. These alignments included sequences from protein-coding and non-coding genes and were used to estimate total substitution rates. For each genome we also created an alignment that only included protein-coding sequences for estimating non-synonymous (d_N) and synonymous (d_S) substitution rates, as well as the ratio of non-synonymous to synonymous substitution rates (d_N/d_S). All non-coding sequences were aligned in Geneious version 6.1.5 (http://www.geneious.com, Kearse et al. 2012) using the MUSCLE plugin (Edgar 2004). The exons of protein-coding genes were aligned in Geneious using the MUSCLE translation alignment plugin. All alignments were then adjusted by eye, using the reading frame and amino acid translations as a guide for protein-coding genes. For each comparison, we only included sequences available for the halophyte, non-halophyte, and outgroup species in the alignments. We also deleted any sites or codons that contained gaps.

Rate estimation and analysis

First we checked that all the collected comparisons had a sufficient number of substitutions to estimate rates by using the test proposed by Welch and Waxman (2008). This test identifies comparisons that are inappropriate for rate analysis due to the stochasticity in substitution rates in shallow comparisons. Comparisons are removed until there is no longer a negative relationship between evolutionary time (total branch length) and contrasts (total differences between halophyte and non-halophyte). Any pairs identified as having an insufficient number of substitutions were removed from the analysis.

Next, we estimated rates for the halophyte and non-halophyte lineages in each comparison. For each comparison, we estimated total substitution rates for the alignments of all sequences in each genome using baseml in the program PAML (Yang 2007) with the REV substitution model and unconstrained rates (clock = 0 in PAML). We estimated d_S, d_N, and d_N/d_S for the protein-coding alignments for each genome in

codeml in PAML (version 4.4b, Yang 2007), using the F3x4 codon frequency model (clock = 0).

Each independent comparison contributed one data point to a non-parametric analysis of the differences in total substitution rate, d_N, d_S and d_N/d_S between halophytes and their non-halophyte relatives. To assess significance, we used the Wilcoxon signed-ranks test (Wilcoxon 1946), which takes into account the sign of the difference in each rate estimate (i.e., non-halophyte rate subtracted from halophyte rate) and the magnitude of the difference in each rate estimate between each halophyte-non-halophyte comparison.

Results

We collected 17 comparisons between mitochondrial sequences from halophytes and their non-halophyte relatives, drawn from 13 orders of angiosperms (Table 1, Table S1). For 15 comparisons we were also able to collect data from the chloroplast genomes, and for 11 we were able to collect sequences from all three genomes (mitochondrial, chloroplast and nuclear sequences). Across the 17 sister comparisons, we found that halophytes had significantly higher total substitution rates than non-halophytes in the mitochondrial sequences (Wilcoxon signed-ranks test, $p = 0.038$, Table 2), taking into account the sign of the difference in rate (i.e., halophyte - non-halophyte) and magnitude of the difference. We did not detect a difference in d_N, d_S, or d_N/d_S in the mitochondrial data. A plot of the data suggests that mitochondrial d_S follows the same pattern as total substitution rates (Figure 1). Although this pattern is not significant, it might indicate low power in our data.

We did not detect any differences in rates of molecular evolution between halophytes and non-halophytes in the chloroplast or nuclear data (Tables 2-4). However, we were only able to collect data from nuclear coding sequences for three comparisons, so we were not able to analyze d_N, d_S, or d_N/d_S in the nuclear data.

Discussion

In this study we used salt tolerant plants, halophytes, as a study system to explore whether exposure to saline habitats increases rates of molecular evolution. We find evidence that halophytes have higher total substitution rates in the mitochondrial sequences analyzed, but not in the chloroplast or nuclear genes. While halophytes had higher total substitution rates in the mitochondrial genes, they did not have higher synonymous or non-synonymous substitution rates in protein-coding regions.

Faster total mitochondrial substitution rates are consistent with the hypothesis that salinity is associated with increased mutation rates due to oxidative damage from ROS on DNA. However, we do not detect an increase in d_S, which is predicted to be proportional to mutation rate (Kimura 1983). For most comparisons, the mitochondrial data consists of only protein-coding genes (Table S2), so estimations of total substitution rate include both non-synonymous changes and synonymous changes over the same alignment length. Including both types of changes increases the power to estimate total changes, which could explain why we detect a significant result in total substitution rate but not in d_N or d_S. Inspection of the data (Figure 1) suggests that the magnitude of the difference in comparisons where the halophyte has a faster mitochondrial d_S is much greater than comparisons where the non-halophyte relative has a faster mitochondrial d_S. Meanwhile the d_N data appears more evenly spread between positive (halophyte faster) and negative (non-halophyte faster) contrasts. This could indicate that including more comparisons or more sequence data might reveal a positive association between halophytes and mitochondrial d_S.

Even if low power is affecting our results, we still find evidence that exposure to salinity is significantly associated with increased total substitution rates in the mitochondrial genome of plants. The effect of environmental factors on rates of molecular evolution has been examined in relatively few case studies (Davies et al. 2004; Whittle 2006; Gillman et al. 2009; Wright et al. 2010; Goldie et al. 2010; Groussin and Gouy 2011; Dowle et al. 2013; Gillman and Wright 2013; Lanfear et al. 2013; Bromham et al. 2015). Thus far studies on the effect of salinity on rates of molecular evolution mostly suggest that salt tolerant organisms experience increased rates (Hebert et al. 2002; Wägele et al. 2003; Baxevanis et al. 2006; Logares et al. 2009,

but see Whittle 2006). Our results are consistent with the prediction that salinity may increase rates of molecular evolution, possibly due to an increase in the production of ROS. Even amongst examples of mutagenic environmental influences, the finding that salt tolerant species have increased rates of molecular evolution is notable. For example, UV is predicted to have a mutagenic affect on DNA as, like salinity, UV can cause damage to cellular structures and induce mutations (Rohde 1992; Britt 1999; Willis et al. 2009). Yet very few comparative studies have explored the relationship between UV exposure (or covariates of UV levels like altitude, Gillman et al. 2009; Wright et al. 2010) and rates of molecular evolution (Dowle et al. 2013; Bromham et al. 2015). We believe that our results represent an intriguing advancement in the study of effects of environmental agents on molecular evolution.

Previous findings suggest that environmental conditions like temperature (Davies et al. 2004; Gillman et al. 2009; Gillman and Wright 2013) and water availability (Goldie et al. 2010) that can increase growth rates may be associated with increased rates of molecular evolution. One explanation for this observation is that increased growth rates lead to a higher rate of genome replications, creating more opportunities for the accumulation of mutations (Lanfear et al. 2013). Although growth in some extremely salt tolerant halophytes is stimulated by moderate to high salinity levels (e.g., Khan et al. 2000; Redondo-Gómez et al. 2006; 2010), a common effect of exposure to salinity is decreased growth rates (Yeo 1983; Munns and Tester 2008). For this reason, increased growth rates are unlikely to explain the observed increase in total substitution rates in mitochondrial genes of halophytes. A more likely alternative hypothesis to ROS damage is that increased rates in halophytes are caused by a physiological strain associated with salinity, for example drought. Drought and salinity impose similar physiological stresses on plants (Munns et al. 2006), and a previous study reported increased molecular rates in drought tolerant plants compared to non-tolerant relatives (Whittle 2006). Therefore, it is possible that the signal we detect is a reflection not of salinity, but of a related environmental factor.

ROS damage is also predicted to affect the chloroplast genome. However, we did not detect any differences between halophytes and non-halophyte relatives in the chloroplast data. As described above, this result could be due to low power. The spread of the data (Table 1) shows that the magnitude of the difference in comparisons where the halophyte has a faster chloroplast d_S tend to be higher than comparisons where the

non-halophyte relative has the higher d_S. However, this pattern is not as strong in the chloroplast data as the mitochondrial data. This pattern suggests that adding more data might reveal that halophytes have consistently faster chloroplast and mitochondrial d_S. Alternatively, an increase in total mitochondrial substitution rates, but not in chloroplast substitution rates, could mean that mitochondria carry a greater stress from exposure to salinity or that mitochondria are particularly important in salinity tolerance (Pastore et al. 2007; Jacoby et al. 2011). ROS are most likely to affect mitochondria endogenously when they are the primary energy producer, during nighttime respiration and in non-photosynthetic tissues like the roots. The roots are often the first and most common point of contact for a plant with soil salinity, so they may accumulate more oxidative damage. The roots also play a critical role in salt tolerance in some halophytes through the active exclusion and selectivity of ions in the soil (Flowers and Colmer 2008). This process is likely to be energetically demanding (Jacoby et al. 2011), potentially exposing root mitochondria to increased stress and thus increased production of ROS. There is also evidence that mitochondrial respiration may become more important in carbon fixation when plants are under water stress (which could be a consequence of drought or salinity), even in photosynthesizing tissues (Flexas et al. 2006). Photosynthetic activity under water stress can decrease due to stomatal closure in response to stress, but mitochondrial respiration remains unaffected or can even increase to supply the chloroplast with more energy (ATP) (Flexas et al. 2006; Atkin and Macherel 2008). This shift in energy demands could lead to increased oxidative damage in the mitochondria. Furthermore, there is evidence to suggest that mitochondria experience more oxidative damage compared to other organelles under stress conditions (Bartoli 2004; Jacoby et al. 2010), which could lead to faster molecular rates in the mitochondrial genome compared to the chloroplast.

Although finding increased substitution rates in mitochondrial genes of halophytes is intriguing, further study is required to elucidate the factors associated with this result and to determine if this pattern is found among a wider sample of halophytes. The data in this analysis was restricted to sequences available on GenBank, which limited the statistical power in terms of finding pairs of halophyte and non-halophyte relatives with matching gene sequences. Many sequences used here originated in studies of broad taxonomic relationships among angiosperms (e.g., Givnish et al. 2011; Qiu et al. 2010; Wurdack and Davis 2009), so many of the halophytes chosen are quite distantly related to the non-halophyte relative. When measuring molecular rates between distantly

related pairs of taxa, many of the substitutions identified would have been acquired deep in the lineage, and so do not necessarily represent changes associated with the development of salt tolerance. Having many comparisons in the data set, as in this study, compensates for the large divergence times of some pairs, though the data set has less statistical power overall for answering the question of whether halophytes have increased molecular rates. Construction of a data set de novo to specifically answer this question would be ideal, which would entail identifying a large range of halophytes and closely related non-halophytes and sequencing several genes from the mitochondrial, nuclear, and chloroplast genomes. Including several distantly related comparisons in this study also reduces the power to identify ecological factors associated with shifts in molecular rates. For example, studies have suggested that several demographic or life history factors may be associated with molecular rates in plants (i.e., generation time, Smith and Donoghue 2008; height, Lanfear et al. 2013). Collecting sequence data specifically for this study would not only allow a more powerful exploration of changes in molecular rates, but also the statistical power to identify life history factors that may be associated with those changes.

Evidence that salinity can influence rates of molecular evolution contributes to the growing literature on the influence of environmental factors (Hebert et al. 2002; Davies et al. 2004; Whittle 2006; Hassanin et al. 2009; Groussin and Gouy 2011; Gillman and Wright 2013) and abiotic stress tolerances (Whittle 2006) on rates of molecular evolution. Specifically, we find that halophytes have increased total substitution rates in the mitochondrial genome, which could be a signal of the mutagenic effect of prolonged or repeated exposure to salinity, possibly caused by increased production of ROS. Our results also suggest that in plants mitochondrial genes may be a useful indicator for future studies of environmental influences on molecular evolution. These results warrant further comparative studies on the general influence of harsh conditions on molecular evolution, including whether environmental influences have different effects on different genomes.

Table 1: Alignment lengths for each halophyte comparison collected for the mitochondrial, chloroplast, and nuclear genomes. See Supplemental Material for details on accession numbers.

Order	Halophyte	Non-halophyte	Mitochondrial		Chloroplast		Nuclear	
			Coding	Total	Coding	Total	Coding	Total
Asparagales	*Asparagus officinalis*	*Hemiphylacus alatostylus*	2613	2613	47808	47808	-	2280
Asterales	*Scaevola plumieri*	*Goodenia ovata*	2571	2571	2625	2983	-	-
Brassicales	*Batis maritima*	*Floerkea proserpinacoides*	4992	4992	7869	7869	-	-
Caryophyllales	*Talinum paniculatum*	*Claytonia virginica*	3717	3717	2031	2678	-	-
Celastrales	*Brexia madagascariensis*	*Plagiopteron suaveolens*	5808	6947	3909	3909	1413	3137
Ericales	*Barringtonia asiatica*	*Couroupita guianensis*	2616	2616	3822	3822	-	1074
-	*Pelliciera rhizophorae*	*Pentamerista neotropica*	2442	2442	4824	4824	-	3296
-	*Planchonella obovata*	*Manilkara zapota*	2661	2661	2661	2661	-	-
-	*Samolus repens*	*Androsace sarmentosa*	2412	2412	-	-	-	-
Laurales	*Cassytha filiformis*	*Laurus nobilis*	4092	4092	2718	2718	-	1884
Malpighiales	*Carallia brachiata*	*Ctenolophon englerianus*	4968	5989	4167	4167	1374	3082
-	*Chrysobalanus icaco*	*Euphronia guianensis*	4197	5607	4224	4224	1359	3046
Malvales	*Althaea officinalis*	*Amoreuxia wrightii*	2652	2652	525	525	-	-
Myrtales	*Terminalia catappa*	*Qualea* sp.	3048	3048	3735	3735	-	1445
Pandanales	*Pandanus tectorius*	*Asplundia rigida*	2730	3826	-	-	-	1687
Piperales	*Anemopsis californica*	*Saururus cernuus*	4029	8360	6543	7365	-	3707
Poales	*Typha latifolia*	*Puya raimondii*	2082	2082	4461	4950	-	1715

Table 2: Estimates of substitution rates for halophytes and non-halophyte comparisons for mitochondrial genes. Values were rounded to three decimal places, so non-zero values appear to be zero and some values appear to be equivalent. Sign indicates whether the halophyte (+) or non-halophyte (-) had a higher value for each rate category. Please see Supplemental Material for further taxonomic information on comparisons, including outgroups. Significant results ($p < 0.05$) for Wilcoxon Signed Ranks tests are indicated with an asterisk.

Order	Halophyte (H)	Non-halophyte (N)	Total			d_N			d_S			d_N/d_S		
			H	N	Sign	H	N	Sign	H	N	Sign	H	N	Sign
Asparagales	*Asparagus officinalis*	*Hemiphylacus alatostylus*	0.002	0.003	-	0.002	0.003	-	0.002	0.005	-	0.670	0.554	+
Asterales	*Scaevola plumieri*	*Goodenia ovata*	0.026	0.017	+	0.002	0.005	-	0.078	0.043	+	0.029	0.120	-
Brassicales	*Batis maritima*	*Floerkea proserpinacoides*	0.017	0.023	-	0.009	0.021	-	0.030	0.025	+	0.296	0.832	-
Caryophyllales	*Talinum paniculatum*	*Claytonia virginica*	0.005	0.003	+	0.002	0.002	-	0.012	0.005	+	0.137	0.377	-
Celastrales	*Brexia madagascariensis*	*Plagiopteron suaveolens*	0.005	0.005	+	0.005	0.003	+	0.007	0.011	-	0.698	0.245	+
Ericales	*Barringtonia asiatica*	*Couroupita guianensis*	0.011	0.007	+	0.010	0.004	+	0.012	0.013	-	0.840	0.306	+
-	*Pelliciera rhizophorae*	*Pentamerista neotropica*	0.011	0.002	+	0.009	0.001	+	0.014	0.002	+	0.687	0.556	+
-	*Planchonella obovata*	*Manilkara zapota*	0.004	0.004	+	0.005	0.005	+	0.002	0.002	+	2.334	2.08	+
-	*Samolus repens*	*Androsace sarmentosa*	0.009	0.006	+	0.007	0.005	+	0.013	0.007	+	0.520	0.685	-
Laurales	*Cassytha filiformis*	*Laurus nobilis*	0.012	0.006	+	0.009	0.005	+	0.017	0.008	+	0.535	0.675	-
Malpighiales	*Carallia brachiata*	*Ctenolophon englerianus*	0.022	0.012	+	0.016	0.010	+	0.026	0.012	+	0.618	0.831	-
-	*Chrysobalanus icaco*	*Euphronia guianensis*	0.010	0.012	-	0.008	0.012	-	0.010	0.011	-	0.753	1.140	-
Malvales	*Althaea officinalis*	*Amoreuxia wrightii*	0.010	0.010	-	0.003	0.002	+	0.025	0.028	-	0.133	0.078	+
Myrtales	*Terminalia catappa*	*Qualea* sp.	0.023	0.014	+	0.011	0.008	+	0.047	0.025	+	0.231	0.331	-
Pandanales	*Pandanus tectorius*	*Asplundia rigida*	0.007	0.007	+	0.007	0.002	+	0.003	0.012	-	1.965	0.149	+
Piperales	*Anemopsis californica*	*Saururus cernuus*	0.004	0.005	-	0.003	0.004	-	0.014	0.014	-	0.237	0.291	-
Poales	*Typha latifolia*	*Puya raimondii*	0.006	0.005	+	0.003	0.001	+	0.014	0.014	-	0.204	0.050	+
	Wilcoxon Signed Ranks Test		$p = 0.045$*			$p = 0.174$			$p = 0.174$			$p = 0.927$		

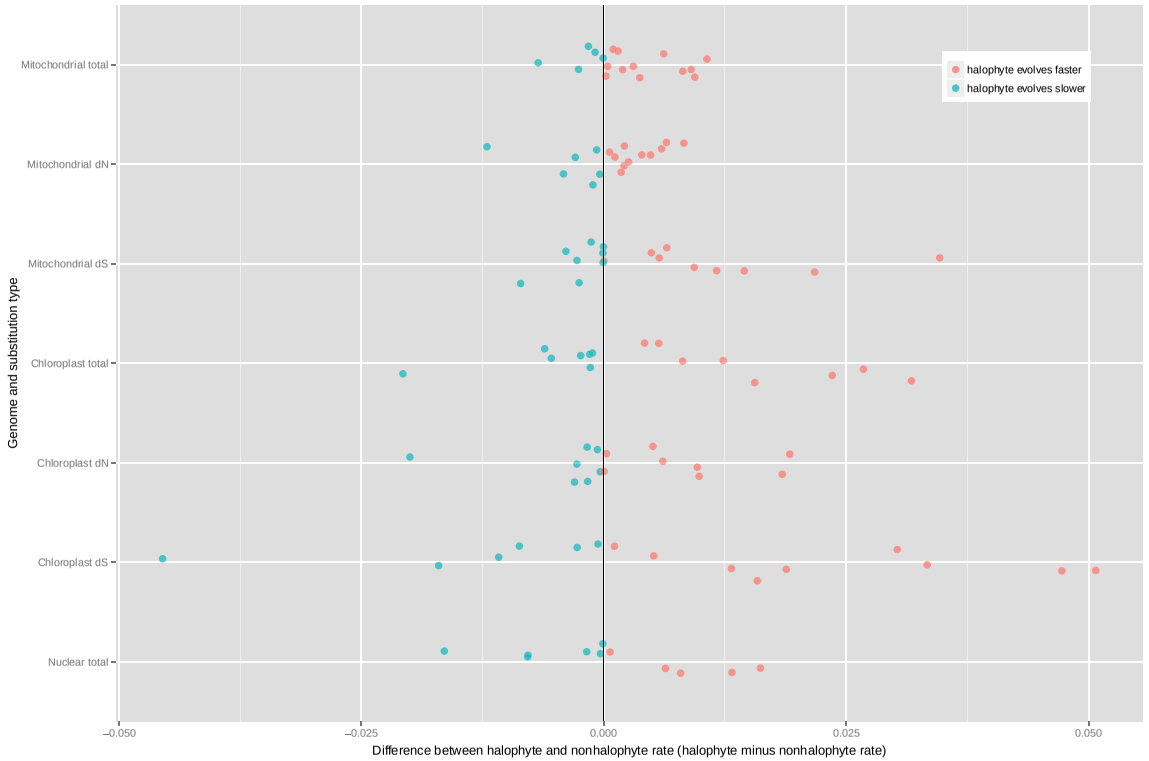
Table 3: Estimates of substitution rates for halophyte comparisons for chloroplast sequences. Values were rounded to three decimal places, so non-zero values appear to be zero and some values appear to be equivalent. Sign indicates whether the halophyte (+) or non-halophyte (-) had a higher value for each rate category. Please see Supplemental Material for further taxonomic information on comparisons, including outgroups.

Order	Halophyte (H)	Non-halophyte (N)	Total			d_N			d_S			d_N/d_S		
			H	N	Sign	H	N	Sign	H	N	Sign	H	N	Sign
Asparagales	*Asparagus officinalis*	*Hemiphylacus alatostylus*	0.005	0.006	-	0.002	0.004	-	0.012	0.013	-	0.177	0.302	-
Asterales	*Scaevola plumieri*	*Goodenia ovata*	0.006	0.012	-	0.003	0.006	-	0.005	0.022	-	0.575	0.258	+
Brassicales	*Batis maritima*	*Floerkea proserpinacoides*	0.083	0.059	+	0.040	0.030	+	0.17	0.119	+	0.236	0.254	-
Caryophyllales	*Talinum paniculatum*	*Claytonia virginica*	0.019	0.039	+	0.007	0.027	-	0.04	0.086	-	0.178	0.317	-
Celastrales	*Brexia madagascariensis*	*Plagiopteron suaveolens*	0.011	0.013	-	0.008	0.008	+	0.016	0.027	-	0.459	0.275	+
Ericales	*Barringtonia asiatica*	*Couroupita guianensis*	0.023	0.019	+	0.009	0.009	+	0.062	0.046	+	0.149	0.195	-
-	*Pelliciera rhizophorae*	*Pentamerista neotropica*	0.027	0.011	+	0.013	0.007	+	0.059	0.025	+	0.22	0.269	-
-	*Planchonella obovata*	*Manilkara zapota*	0.004	0.005	-	0.002	0.004	-	0.008	0.007	+	0.303	0.596	-
Laurales	*Cassytha filiformis*	*Laurus nobilis*	0.058	0.026	+	0.035	0.017	+	0.093	0.046	+	0.38	0.369	+
Malpighiales	*Carallia brachiata*	*Ctenolophon englerianus*	0.050	0.044	-	0.028	0.023	+	0.105	0.100	+	0.262	0.225	+
-	*Chrysobalanus icaco*	*Euphronia guianensis*	0.040	0.046	+	0.022	0.025	-	0.088	0.097	-	0.245	0.254	-
Malvales	*Althaea officinalis*	*Amoreuxia wrightii*	0.040	0.013	-	0.019	0.000	+	0.074	0.055	+	0.26	0.000	+
Myrtales	*Terminalia catappa*	*Qualea* sp.	0.056	0.048	-	0.024	0.025	-	0.124	0.094	+	0.195	0.266	-
Piperales	*Anemopsis californica*	*Saururus cernuus*	0.011	0.012	+	0.006	0.007	-	0.022	0.025	-	0.288	0.271	+
Poales	*Typha latifolia*	*Puya raimondii*	0.039	0.026	+	0.021	0.012	+	0.079	0.066	+	0.268	0.175	+
	Wilcoxon Signed Ranks Test, all comparisons		$p = 0.188$			$p = 0.524$			$p = 0.208$			$p = 0.978$		

Table 4: Estimates of substitution rates for halophyte comparisons for nuclear sequences. Values were rounded to three decimal places, so non-zero values appear to be zero and some values appear to be equivalent. Sign indicates whether the halophyte (+) or non-halophyte (-) had a higher value for each rate category. Please see Supplemental Material for further taxonomic information on comparisons, including outgroups.

Order	Halophyte (H)	Non-halophyte (N)	Total			d_N			d_S			d_N/d_S		
			H	N	Sign	H	N	Sign	H	N	Sign	H	N	Sign
Asparagales	*Asparagus officinalis*	*Hemiphylacus alatostylus*	0.004	0.006	-									
Celastrales	*Brexia madagascariensis*	*Plagiopteron suaveolens*	0.018	0.017	+	0.009	0.009	+	0.088	0.092	-	0.104	0.093	+
Ericales	*Barringtonia asiatica*	*Couroupita guianensis*	0.005	0.022	-									
-	*Pelliciera rhizophorae*	*Pentamerista neotropica*	0.009	0.010	-									
Laurales	*Cassytha filiformis*	*Laurus nobilis*	0.019	0.006	+									
Malpighiales	*Carallia brachiata*	*Ctenolophon englerianus*	0.037	0.030	+	0.021	0.021	+	0.19	0.16	+	0.113	0.129	-
-	*Chrysobalanus icaco*	*Euphronia guianensis*	0.040	0.024	+	0.024	0.021		0.148	0.114	+	0.161	0.186	-
Myrtales	*Terminalia catappa*	*Qualea* sp.	0.011	0.019	-									
Pandanales	*Pandanus tectorius*	*Asplundia rigida*	0.007	0.007	-									
Piperales	*Anemopsis californica*	*Saururus cernuus*	0.014	0.006	+									
Poales	*Typha latifolia*	*Puya raimondii*	0.017	0.025	-									
		Wilcoxon Signed Ranks Test, all comparisons	$p = 0.898$											

Figure 1: Scatter plot of the difference in rates between halophytes and non-halophytes. Rate differences were calculated by subtracting the non-halophyte rate from the halophyte rate. Points where the halophyte had a faster rate are positive (red dots) and points where the non-halophyte had a faster rate are negative (blue dots). Each line on the y-axis represents a different type of molecular rate estimated in the analysis (see Methods). The range in differences was much larger for d_N/d_S estimates, so we have excluded them from the plot. Points have been jittered vertically so that each point can be seen clearly.



References

Apel, K., and H. Hirt. 2004. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annual Review of Plant Biology 55:373–399.

Atkin, O. K., and D. Macherel. 2008. The crucial role of plant mitochondria in orchestrating drought tolerance. Annals of Botany 103:581–597.

Bartoli, C. G. 2004. Mitochondria are the main target for oxidative damage in leaves of wheat (*Triticum aestivum* L.). Journal of Experimental Botany 55:1663–1669.

Baxevanis, A. D., I. Kappas, and T. J. Abatzopoulos. 2006. Molecular phylogenetics and asexuality in the brine shrimp *Artemia*. Molecular Phylogenetics and Evolution 40:724–738.

Bennett, T. H., T. J. Flowers, and L. Bromham. 2013. Repeated evolution of salt-tolerance in grasses. Biology letters 9:20130029–20130029.

Bose, J., A. Rodrigo-Moreno, and S. Shabala. 2014. ROS homeostasis in halophytes in the context of salinity stress tolerance. Journal of Experimental Botany 65:1241–1257.

Britt, A. 1999. Molecular genetics of DNA repair in higher plants. Trends in Plant Science 4:20–25.

Bromham, L., X. Hua, R. Lanfear, and P. F. Cowman. 2015. Exploring the relationships between mutation rates, life history, genome size, environment, and species richness in flowering plants. The American Naturalist. DOI:10.1086/680052

Cai-Hong, P., Z. Su-Jun, G. Zhi-Zhong, and W. Bao-Shan. 2005. NaCl treatment markedly enhances H_2O_2-scavenging system in leaves of halophyte *Suaeda salsa*. Physiologia Plantarum 125:490–499.

Chamberlain, S., and E. Szocs. 2013. taxize - taxonomic search and retrieval in R. F1000 Research. 2:191. http://f1000research.com/articles/2-191/v2.

Chamberlain, S., E. Szocs, C. Boettiger, K. Ram, I. Bartomeus, and J. Baumgartner. 2014. taxize: Taxonomic information from around the web. R package version 0.3.0. https://github.com/ ropensci/taxize.

Davies, T. J., V. Savolainen, M. W. Chase, J. Moat, and T. G. Barraclough. 2004. Environmental energy and evolutionary rates in flowering plants. Proceedings of the Royal Society of London B: Biological Sciences 271:2195–2200.

Dennis, P. P., and L. C. Shimmin. 1997. Evolutionary divergence and salinity-mediated selection in halophilic Archaea. Microbiology and Molecular Biology Reviews 61:90–104.

Dowle, E. J., M. Morgan-Richards, and S. A. Trewick. 2013. Molecular evolution and the latitudinal biodiversity gradient. Heredity 110:501–510.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792–1797.

Ellouzi, H., K. Ben Hamed, J. Cela, S. Munné-Bosch, and C. Abdelly. 2011. Early effects of salt stress on the physiological and oxidative status of *Cakile maritima* (halophyte) and *Arabidopsis thaliana* (glycophyte). Physiologia Plantarum 142:128–143.

Flexas, J., J. Bota, J. Galmes, H. Medrano, and M. Ribas Carbó. 2006. Keeping a positive carbon balance under adverse conditions: responses of photosynthesis and respiration to water stress. Physiologia Plantarum 127:343–352.

Flowers, T. J., and T. D. Colmer. 2008. Salinity tolerance in halophytes. New Phytologist 179:945–963.

Flowers, T. J., H. K. Galal, and L. Bromham. 2010. Evolution of halophytes: multiple origins of salt tolerance in land plants. Functional Plant Biology 37:604.

Flowers, T., P. F. Troke, and A. R. Yeo. 1977. The mechanism of salt tolerance in halophytes. Annual Review of Plant Physiology 28:89–121.

Foyer, C. H., and G. Noctor. 2005. Redox homeostasis and antioxidant signaling: a metabolic interface between stress perception and physiological responses. The Plant Cell 17:1866–1875.

Fujita, M., Y. Fujita, Y. Noutoshi, F. Takahashi, Y. Narusaka, K. Yamaguchi-Shinozaki, and K. Shinozaki. 2006. Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. Current Opinion in Plant Biology 9:436–442.

Galtier, N. 2011. The intriguing evolutionary dynamics of plant mitochondrial DNA. BMC Biology 9:61.

Gill, S. S., and N. Tuteja. 2010. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. Plant Physiology and Biochemistry. 48:909–930.

Gillman, L. N., and S. D. Wright. 2013. Species richness and evolutionary speed: the influence of temperature, water and area. Journal of Biogeography 41:39–51.

Gillman, L. N., D. J. Keeling, H. A. Ross, and S. D. Wright. 2009. Latitude, elevation and the tempo of molecular evolution in mammals. Proceedings of the Royal Society of London B: Biological Sciences 276:3353–3359.

Givnish, T. J., M. H. J. Barfuss, B. Van Ee, R. Riina, K. Schulte, R. Horres, et al. 2011. Phylogeny, adaptive radiation, and historical biogeography in Bromeliaceae: Insights from an eight-locus plastic phylogeny. American Journal of Botany 98:872–895.

Glenn, E. P., J. J. Brown, and E. Blumwald. 1999. Salt tolerance and crop potential of halophytes. Critical Reviews in Plant Science 18:227–255.

Goldie, X., L. Gillman, M. Crisp, and S. Wright. 2010. Evolutionary speed limited by water in arid Australia. Proceedings of the Royal Society of London B: Biological Sciences 277:2645–2653.

Gong, X., L. Chao, M. Zhou, M. Hong, L. Luo, L. Wang, W. Ying, C. Jingwei, G. Songjie, and H. Fashui. 2010. Oxidative damages of maize seedlings caused by exposure to a combination of potassium deficiency and salt stress. Plant Soil 340:443–

452.

Groussin, M., and M. Gouy. 2011. Adaptation to environmental temperature is a major determinant of molecular evolutionary rates in Archaea. Molecular Biology and Evolution 28:2661–2674.

Halliwell, B. 2006. Reactive Species and Antioxidants. Redox biology is a fundamental theme of aerobic life. Plant Physiology 141:312–322.

Hassanin, A., A. Ropiquet, A. Couloux, and C. Cruaud. 2009. Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). Journal of Molecular Evolution 68:293–310.

Hebert, P. D. N., E. A. Remigio, J. K. Colbourne, D. J. Taylor, and C. C. Wilson. 2002. Accelerated molecular evolution in halophilic crustaceans. Evolution 56:909–926.

Jacoby, R. P., A. H. Millar, and N. L. Taylor. 2010. Wheat mitochondrial proteomes provide new links between antioxidant defense and plant salinity tolerance. Journal of Proteome Research 9:6595–6604.

Jacoby, R. P., N. L. Taylor, and A. H. Millar. 2011. The role of mitochondrial respiration insalinity tolerance. Trends in Plant Science 16:614–623.

Jiang, C., A. Mithani, E. J. Belfield, R. Mott, L. D. Hurst, and N. P. Harberd. 2014. Environmentally responsive genome-wide accumulation of de novo Arabidopsis thalianamutations and epimutations. Genome Research 24:1821–1829.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Khan, M. A., I. A. Ungar, and A. M. Showalter. 2000. Effects of salinity on growth, water relations and ion accumulation of the subtropical perennial halophyte, *Atriplex griffithii* var. *stocksii*. Annals of Botany 85:225–232.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

Klekowski, E. J., and P. J. Godfrey. 1989. Ageing and mutation in plants. Nature 340:389–391.

Klekowski, E. J., Jr. 1988. Genetic load and its causes in long-lived plants. Trees 2:195–203.

Knoop, V. 2004. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Current Genetics 46:123–139.

Lanfear, R., J. J. Welch, and L. Bromham. 2010. Watching the clock: Studying variation in rates of molecular evolution between species. Trends in Ecology & Evolution 25:495–503.

Lanfear, R., S. Y. W. Ho, T. Jonathan Davies, A. T. Moles, L. Aarssen, N. G. Swenson,

L. Warman, A. E. Zanne, and A. P. Allen. 2013. Taller plants have lower rates of molecular evolution. Nature Communications 4:1879.

Lechno, S., E. Zamski, and E. Tel-Or. 1997. Salt stress-induced responses. Journal of Plant Physiology 150:206–211.

Logares, R., J. Brate, F. Heinrich, K. Shalchian-Tabrizi, and S. Bertilsson. 2010. Infrequent transitions between saline and fresh waters in one of the most abundant microbial lineages (SAR11). Molecular Biology and Evolution 27:347–357.

Logares, R., J. Bråte, S. Bertilsson, J. L. Clasen, K. Shalchian-Tabrizi, and K. Rengefors. 2009. Infrequent marine–freshwater transitions in the microbial world. Trends in Microbiology 17:414–422.

Lushchak, V. I. 2011. Environmentally induced oxidative stress in aquatic animals. Aquatic Toxicology 101:13–30.

Lynch, M. 2010. Evolution of the mutation rate. Trends in Genetics 26:345–352.

Martínez-Alvarez, R. M., M. C. Hidalgo, A. Domezain, A. E. Morales, M. García-Gallego, and A. Sanz. 2002. Physiological changes of sturgeon *Acipenser naccarii* caused by increasing environmental salinity. J. Exp. Biol. 205:3699–3706.

Miller, G., V. Shulaev, and R. Mittler. 2008. Reactive oxygen signaling and abiotic stress. Physiologia Plantarum 133:481–489.

Mittler, R. 2002. Oxidative stress, antioxidants and stress tolerance. Trends in Plant Science 7:405–410.

Mittler, R., S. Vanderauwera, M. Gollery, and F. Van Breusegem. 2004. Reactive oxygen gene network of plants. Trends in Plant Science 9:490–498.

Mittler, R., S. Vanderauwera, N. Suzuki, G. Miller, V. B. Tognetti, K. Vandepoele, M. Gollery, V. Shulaev, and F. Van Breusegem. 2011. ROS signaling: the new wave? Trends in Plant Science 16:300–309.

Møller, I. M. 2001. Plant mitochondria and oxidative stress: Electron transport, NADPH turnover, and metabolism of reactive oxygen species. Annual Review of Plant Physiology and Plant Molecular Biology 52:561–591.

Munns, R., and M. Tester. 2008. Mechanisms of salinity tolerance. Annual Review of Plant Biology 59:651–681.

Munns, R., R. A. James, and A. Läuchli. 2006. Approaches to increasing the salt tolerance of wheat and other cereals. Journal of Experimental Botany 57:1025–1043.

Navrot, N., N. Rouhier, E. Gelhaye, and J.-P. Jacquot. 2007. Reactive oxygen species generation and antioxidant systems in plant mitochondria. Physiologia Plantarum 129:185–195.

Ozgur, R., B. Uzilday, A. H. Sekmen, and I. Turkan. 2013. Reactive oxygen species regulation and antioxidant defence in halophytes. Functional Plant Biology 40:832–847.

Parida, A. K., and A. B. Das. 2005. Salt tolerance and salinity effects on plants: a

review. Ecotoxicology and Environmental Safety 60:324–349.

Parida, A. K., and B. Jha. 2009. Antioxidative defense potential to salinity in the euhalophyte *Salicornia brachiata*. Journal of Plant Growth and Regulation 29:137–148.

Pastore, D., D. Trono, M. N. Laus, N. Di Fonzo, and Z. Flagella. 2007. Possible plant mitochondria involvement in cell adaptation to drought stress. A case study: durum wheat mitochondria. Journal of Experimental Botany 58:195–210.

Qiu, Y.-L., L. Libo, J.-Y. Xue, T. A. Hendry, R.-Q. Li, J. W. Brown, Y. Liu, G. T. Hudson, and Z.-D. Chen. 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. Journal of Systematics and Evolution. 48:391–425.

Redondo-Gómez, S., C. Wharmby, J. M. Castillo, E. Mateos-Naranjo, C. J. Luque, A. de Cires, T. Luque, A. J. Davy, and M. Enrique Figueroa. 2006. Growth and photosynthetic responses to salinity in an extreme halophyte, *Sarcocornia fruticosa*. Physiologia Plantarum 128:116–124.

Redondo-Gómez, S., E. Mateos-Naranjo, M. E. Figueroa, and A. J. Davy. 2010. Salt stimulation of growth and photosynthesis in an extreme halophyte, *Arthrocnemum macrostachyum*. Plant Biology 12:79–87.

Rhoads, D. M. 2006. Mitochondrial reactive oxygen species. Contribution to oxidative stress and interorganellar signaling. Plant Physiology 141:357–366.

Rohde, K. 1992. Latitudinal gradients in species diversity: the search for the primary cause. Oikos 65:514–527.

Roldán-Arjona, T., and R. R. Ariza. 2009. Repair and tolerance of oxidative DNA damage in plants. Mutation Research/Reviews in Mutation Research 681:169–179.

Seckin, B., I. Turkan, A. H. Sekmen, and C. Ozfidan. 2010. The role of antioxidant defense systems at differential salt tolerance of *Hordeum marinum* Huds. (sea barleygrass) and *Hordeum vulgare* L. (cultivated barley). Environmental and Experimental Botany 69:76–85.

Slama, I., C. Abdelly, A. Bouchereau, T. Flowers, and A. Savoure. 2015. Diversity, distribution and roles of osmoprotective compounds accumulated in halophytes under abiotic stress. Annals of Botany 115:433–447.

Smith, A. S., and M. J. Donoghue. 2008. Rates of molecular evolution are linked to life history in flowering plants. Science 322:86–89.

Song, F.-N., C.-P. Yang, X.-M. Liu, and G.-B. Li. 2006. Effect of salt stress on activity of superoxide dismutase (SOD) in *Ulmus pumila* L. Journal of Forestry Research 17:13–16. Springer.

Stewart, G. R., and J. A. Lee. 1974. The role of proline accumulation in halophytes. Planta 120:279–289.

The Plant List. 2010. The Plant List. Version 2. http://theplantlist.org.

Uzilday, B., R. Ozgur, A. H. Sekmen, E. Yildiztugay, and I. Turkan. 2015. Changes in the alternative electron sinks and antioxidant defense in chloroplasts of the extreme

halophyte *Eutrema parvulum* (*Thellungiella parvula*) under salinity. Annals of Botany 115:449–463.

Wägele, J.-W., B. Holland, H. Dreyer, and B. Hackethal. 2003. Searching factors causing implausible non-monophyly: ssu rDNA phylogeny of Isopoda Asellota (Crustacea: Peracarida) and faster evolution in marine than in freshwater habitats. Molecular Phylogenetics and Evolution 28:536–551.

Welch, J. J., and D. Waxman. 2008. Calculating independent contrasts for the comparative study of substitution rates. Journal of Theoretical Biology 251:667–678.

Whittle, C. A. 2006. The influence of environmental factors, the pollen : ovule ratio and seed bank persistence on molecular evolutionary rates in plants. Journal of Evolutionary Biology 19:302–308.

Wilcoxon, F. 1946. Individual comparisons of grouped data by ranking methods. Journal of Economic Entomology 39:269.

Willis, K. J., K. D. Bennett, and H. J. B. Birks. 2009. Variability in thermal and UV-B energy fluxes through time and their influence on plant diversity and speciation. Journal of Biogeography 36:1630–1644.

Wolfe, K. H., W.-H. Li, and P. M. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Sciences of the United States of America 84:9054–9058.

Wright, S. D., H. A. Ross, D. Jeanette Keeling, P. McBride, and L. N. Gillman. 2010. Thermal energy and the rate of genetic evolution in marine fishes. Evolutionary Ecology 25:525–530.

Wurdack, K. J., and C. C. Davis. 2009. Malpighiales phylogenetics: Gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. American Journal of Botany 96:1551–1570.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24:1586–1591.

Yeo, A. R. 1983. Salinity resistance: Physiologies and prices. Physiologia Plantarum 58:214–222.

Supplementary Material

Contents:

Table S1: Species included in each halophyte comparison. Each row represents a separate comparison. See Methods for selection criteria of halophytes and non-halophytes. Taxonomy follows The Plant List (2010).

Order (H)	Family (H)	Halophyte	Order (N)	Family (N)	Non-halophyte	Order (O)	Family (O)	Outgroup
Asparagales	Asparagaceae	*Asparagus officinalis*	Asparagales	Asparagaceae	*Hemiphylacus alatostylus*	Asparagales	Asparagaceae	*Sansevieria trifasciata*
Asterales	Goodeniaceae	*Scaevola plumieri*	Asterales	Goodeniaceae	*Goodenia ovata*	Asterales	Asteraceae	*Helianthus annuus*
Brassicales	Bataceae	*Batis maritima*	Brassicales	Limnanthaceae	*Floerkea proserpinacoides*	Malvales	Malvaceae	*Gossypium arboreum*
Caryophyllales	Talinaceae	*Talinum paniculatum*	Caryophyllales	Montiaceae	*Claytonia virginica*	Caryophyllales	Limeaceae	*Limeum africanum*
Celastrales	Celastraceae	*Brexia madagascariensis*	Celastrales	Celastraceae	*Plagiopteron suaveolens*	Celastrales	Celastraceae	*Stackhousia minima*
Ericales	Lecythidaceae	*Barringtonia asiatica*	Ericales	Lecythidaceae	*Couroupita guianensis*	Ericales	Polemoniaceae	*Polemonium caeruleum*
-	Tetrameristaceae	*Pelliciera rhizophorae*	-	Tetrameristaceae	*Pentamerista neotropica*	-	Fouquieriaceae	*Fouquieria columnaris*
-	Sapotaceae	*Planchonella obovata*	-	Sapotaceae	*Manilkara zapota*	-	Lecythidaceae	*Couroupita guianensis*
-	Primulaceae	*Samolus repens*	-	Primulaceae	*Androsace sarmentosa*	-	Sapotaceae	*Manilkara zapota*
Laurales	Lauraceae	*Cassytha filiformis*	Laurales	Lauraceae	*Laurus nobilis*	Laurales	Calycanthaceae	*Idiospermum australiense*
Malpighiales	Rhizophoraceae	*Carallia brachiata*	Malpighiales	Ctenolophonaceae	*Ctenolophon englerianus*	Malpighiales	Malpighiaceae	*Dicella nucifera*
-	Chrysobalanaceae	*Chrysobalanus icaco*	-	Euphroniaceae	*Euphronia guianensis*	-	Euphorbiaceae	*Moultonianthus leembruggianus*
Malvales	Malvaceae	*Althaea officinalis*	Malvales	Bixaceae	*Amoreuxia wrightii*	Brassicales	Brassicaceae	*Arabidopsis thaliana*
Myrtales	Combretaceae	*Terminalia catappa*	Myrtales	Vochysiaceae	*Qualea* sp	Myrtales	Penaeaceae	*Olinia emarginata*
Pandanales	Pandanaceae	*Pandanus tectorius*	Pandanales	Cyclanthaceae	*Asplundia rigida*	Pandanales	Triuridaceae	*Kupea martinetugei*
Piperales	Saururaceae	*Anemopsis californica*	Piperales	Saururaceae	*Saururus cernuus*	Piperales	Aristolochiaceae	*Aristolochia macrophylla*
Poales	Typhaceae	*Typha latifolia*	Poales	Bromeliaceae	*Puya raimondii*	Commelinales	Pontederiaceae	*Pontederia cordata*

Table S2: GenBank accessions for genes included in analysis. Sequences for *Asparagus officinalis*, *Hemiphylacus alatostylus*, and *Sansevieria trifasciata* are not included, as we collected a large amount of sequences for this comparison. We collected sequences from Steele *et al*. (2012) that were available for all three species in the comparison, including 66 protein-coding chloroplast genes, 5 protein-coding mitochondrial genes, and 2 non-coding nuclear genes. The type column designates if the species is the halophyte (H), non-halophyte (N), or outgroup (O) in the comparison.

					Mitochondrial genes											Chloroplast genes							Nuclear genes			
Pair	Order	Family	Species	Type	matR	atp1	atpA	cob	ccmb	nad1	nad5	nad6	cox1	rps3	18s	26s	rbcL	matK	atpB	ndhF	rpoC2	trnK	trnL-tmF	18S	26S	EMB2765	phyC
1	Asparagales	Asparagaceae	*Asparagus officinalis*	H																							
1			*Hemiphylacus alatostylus*	N																							
1			*Sansevieria trifasciata*	O																							
2	Asterales	Goodeniaceae	*Scaevola plumieri*	H		EU280989							EU281064				EU841126	JQ711557					JQ682695				
2			*Goodenia ovata*	N		AY818934							AY053575				X87386	JQ711602					JQ682776				
2		Asteraceae	*Helianthus annuus*	O		NC023337							NC023337				L13929	AY215805					DQ131560				
3	Brassicales	Bataceae	*Batis maritima*	H	GU351158	GU350968					GU351346			GU351578			M88341	AY483219	AF209538	EU002199	EU002474						
3		Limnanthaceae	*Floerkea proserpinacoides*	N	GU351210	GU351021					GU351407			GU351663			L12679	EU002178	AF035904	JF268472	EU002502						
3	Malvales	Malvaceae	*Gossypium arboreum*	O	GU351218	GU351029					GU351420			GU351676			GU981720	NC016712	GU981695	U55331	NC016712						
4	Caryophyllales	Portulacaceae	*Talinum paniculatum*	H	GU351313	GU351124					GU351526						HM850388	AY015274		AF194830		AY015274					
4			*Claytonia virginica*	N	GU351184	GU350992					GU351375							AY764113		AF194856		AY764113			HQ843445		
4		Limeaceae	*Limeum africanum*	O	GU351241	GU351051					GU351445							JQ844143		HQ843272		JQ844143			HQ843452		
5	Celastrales	Celastraceae	*Brexia madagascariensis*	H	AY674482			EF135193	EF135086	AY674655		EF135321		EF135422			L11176	EF135510	AJ235419					U42543		FJ669851	FJ669950
5			*Plagiopteron suaveolens*	N	FJ670399			FJ670332	FJ670250	AY674751		FJ669562		FJ669664			FJ670182	FJ670052	AJ235562					AF206993		FJ669854	FJ669952
5			*Stackhousia minima*	O	AY674564			EF135276	EF135170	AY674773		FJ669564		EF135487			FJ670184	EF135596	AJ235610					AF207026		FJ669856	AY425131
6	Ericales	Lecythidaceae	*Barringtonia asiatica*	H	AF420992	AY725906											EU980812		DQ923986	AF421044			DQ417930		AY727949		
6			*Couroupita guianensis*	N	GU351190	AY725907											AF022134		AJ236224	AF421050			DQ417951		AY727950		
6		Polemoniaceae	*Polemonium caeruleum*	O	AF421023	AF420944											KF997406		AF420980	AF421070			EU348478		AY056516		
7		Tetrameristaceae	*Pelliciera rhizophorae*	H	AF421022	AY725900											AF206804		AF209647	AY725871					AY727933		
7			*Pentamerista neotropica*	N	AY725886	AY725901											AY725860		AY725921	AY725872					AY727935		
7		Fouquieriaceae	*Fouquieria columnaris*	O	GU351234	AY725902											AY725861		AJ235501	AF207961					AF479159		
8		Primulaceae	*Samolus repens*	H	AF421027	AF420946																					
8			*Andrasace sarmentosa*	N	GU351149	GU350959																					
8		Sapotaceae	*Manilkara zapota*	O	AF421016	AF420938																					
9			*Planchonella obovata*	H	AF421024	AF420945											AF421100		AF420981	AF421071			AB817474	EU661402	AB114899		
9			*Manilkara zapota*	N	AF421016	AF420938											EU980807		AJ235528	AF213732			DQ924309	U43080	AF479160		
9		Lecythidaceae	*Couroupita guianensis*	O	GU351190	AY725907											AF022134		AJ236224	AF421050			DQ417951	AJ235993	AY727950		
10	Laurales	Lauraceae	*Cassytha filiformis*	H	EU281123	EU281002							EU281076				JQ933256	GU117739						KC588400			
10			*Laurus nobilis*	N	AF197798	AF197682							AF193956				AY337731	AF244407						AF197580			
10		Calycanthaceae	*Idiospermum australiense*	O	AF197779	AF197680							DQ508956				L12651	AY525342						AF206937			
11	Malpighiales	Chrysobalanaceae	*Chrysobalanus icaco*	H	AY674491			EF135202	EF135093	AY674668		EF135331		EF135431			L11178	EF135519	AF209562	FJ670067				U42519		FJ669737	FJ669885
11		Euphroniaceae	*Euphronia guianensis*	N	AY674513			EF135220	EF135115	AY674696		EF135348		FJ669610			AF089762	EF135540	AY788221	AY425052				AY674597		FJ669773	FJ669904
11			*Moultonianthus leembruggianus*	O	FJ670366			FJ670289	FJ670207	FJ670423		FJ669513		FJ669602			AY794982	FJ670015	FJ669974	FJ670086				FJ669695		FJ669762	FJ669899
12		Rhizophoraceae	*Carallia brachiata*	H	AY674485			EF135197	FJ670235	AY674660		EF135325		EF135425			AF206744	EF135513	AJ235425	AY425038				FJ707525		FJ669826	AY674448
12		Ctenolophonaceae	*Ctenolophon englerianus*	N	AY674499			EF135207	EF135098	AY674676		EF135335		EF135436			AJ402940	EF135524	AY788215	FJ670074				AY674589		FJ669745	FJ669892
12		Malpighiaceae	*Dicella nucifera*	O	AY674501			EF135210	EF135101	AY674681		EF135337		EF135437			AJ235802	AF344541	AJ235453	AF351048				AF206901		FJ669793	AF500553
13	Malvales	Malvaceae	*Althaea officinalis*	H		EU280970							EU281042				JN891502										
13		Bixaceae	*Amoreuxia wrightii*	N		EU280957							EU281028				Y15142										
13	Brassicales	Brassicaceae	*Arabidopsis thaliana*	O		NC001284							NC001284				AB917053										
14	Myrtales	Combretaceae	*Terminalia catappa*	H		GU351127					GU351529			GU351833			FJ381811	GU135057	AF209686	EU002267				AF207037			
14		Vochysiaceae	*Qualea* sp.	N		GU351096					GU351496			GU351784			U02730	AF368216	AF209662	AY498829				AF207002			
14		Penaeaceae	*Olinia emarginata*	O		GU351074					GU351473			GU351752			AJ605089	JX970901	GQ497640	AJ605102				AM235487			
15	Pandanales	Pandanaceae	*Pandanus tectorius*	H	KF298331		KF298317			KF298345														KF298352			
15		Cyclanthaceae	*Asplundia rigida*	N	KF258578		KF258199			KF258207														KF264477			
15		Stemonaceae	*Kupea martinetugei*	O	KF197110		GQ469520			KF197103														KF197093			
16	Piperales	Saururaceae	*Anemopsis californica*	H	AF197747	AF197631					DQ406979				DQ008730	DQ008791	AY572258	DQ882198	AF197608		DQ407110	DQ882198		AF197576	DQ008639		
16			*Saururus cernuus*	N	AF197748	AF197633					DQ406934				DQ008732	DQ008792	L14294	DQ882200	AF093398		DQ407094	DQ882200		U42805	AY095468		
16		Aristolochiaceae	*Aristolochia macrophylla*	O	AF197732	AF197669					DQ406911				DQ008674	DQ008796	L12630	AB060742	AF528845		DQ407095	DQ882193		AF206855	AY095450		
17	Poales	Typhaceae	*Typha latifolia*	H		AY124510		DQ916721									DQ069503	AB088801	AF168951	NC013823			HQ882766	AF168880			
17		Bromeliaceae	*Puya raimondii*	N		AY124508		DQ916703									AF206814	EU780852	AF209661	AY438611			HQ882754	AF207001			
17	Commelinales	Pontederiaceae	*Pontederia cordata*	O		EU280987		DQ916698									U41594	AF434872	AF209657	U41618			AJ387743	AF206998			

References

Steele, P.R., K. L. Hertweck, D. Mayfield, M. R. McKain, J. Leebens-Mack, and J. C. Pires. 2012. Quality and quantity of data recovered from massively parallel sequencing: Examples in Asparagales and Poaceae. American Journal of Botany 99: 330–348.

The Plant List. 2010. The Plant List. Version 2. http://theplantlist.org.

Chapter 5

Domestication and the mitochondrial genome: comparing patterns and rates of molecular evolution in domesticated mammals and birds and their wild relatives

Camile Moray, Robert Lanfear, and Lindell Bromham

Published: Genome Biology and Evolution. 2014. 6(1): 161–169

Abstract

Studies of domesticated animals have led to the suggestion that domestication could have significant effects on patterns of molecular evolution. In particular, analyses of mitochondrial genome sequences from domestic dogs and yaks have yielded higher ratios of non-synonymous to synonymous substitutions in the domesticated lineages than in their wild relatives. These results are important because they imply that changes to selection or population size operating over a short timescale can cause significant changes to the patterns of mitochondrial molecular evolution. In this study, our aim is to test whether the impact on mitochondrial genome evolution is a general feature of domestication, or whether it is specific to particular examples. We test whether domesticated mammals and birds have consistently different patterns of molecular evolution than their wild relatives for 16 phylogenetically independent comparisons of mitochondrial genome sequences. We find no consistent difference in branch lengths or d_N/d_S between domesticated and wild lineages. We also find no evidence that our failure to detect a consistent pattern is due to the short timescales involved, or low genetic distance between domesticated lineages and their wild relatives. However, removing comparisons where the wild relative may also have undergone a bottleneck does reveal a pattern consistent with reduced effective population size in domesticated lineages. Our results suggest that, while some domesticated lineages may have undergone changes to selective regime or effective population size that could have affected mitochondrial evolution, it is not possible to generalise these patterns over all domesticated mammals and birds.

Key Words: relaxed selection, artificial selection, mitochondria, d_N/d_S, effective population size, comparative analysis

Introduction

Does domestication influence rates and patterns of molecular evolution? Analysis of SNPs from the dog nuclear genome suggests a higher ratio of non-synonymous to synonymous alleles relative to the wolf, which has been interpreted as the signature of relaxed selection and reduction in effective population size associated with domestication (Cruz et al. 2008). Similarly, studies have found that rice (Lu et al. 2006) and a laboratory strain of yeast (Gu et al. 2005) have higher ratios of non-synonymous to synonymous changes (d_N/d_S) than their wild relatives. Comparison of dog, yak, pig and silkworm mitochondrial genomes to their respective wild relatives have also shown that the domesticated lineages have higher d_N/d_S than their wild relatives (Björnerfeldt et al. 2006; Wang et al. 2011; Hughes 2013).

These studies raise the possibility that domestication has significant effects on molecular evolution. If true, this would demonstrate that rates and patterns of molecular evolution are labile on relatively short timescales. It is widely assumed that all domesticated lineages were established less than 15,000 years ago, so any detectable effects of domestication on molecular evolution must be due to recent changes having a significant and measurable impact on molecular evolution. Domesticated lineages might therefore provide an interesting case study for the influence of population changes or alteration of selective regime on patterns and rates of molecular evolution. On a practical level, observation of widespread impacts of domestication on molecular evolution would suggest that caution must be exercised when estimating the date of origin of domesticated lineages from molecular data, or when including sequences from domesticated lineages in dating analyses.

Broadly speaking, there are three ways that domestication could affect patterns of molecular evolution: artificial selection, relaxed selective constraints, and reduced effective population size in domesticated lineages. Direct or indirect selection for traits during domestication may increase the rate of non-synonymous substitutions at specific loci associated with selected traits (e.g., coat colour in pigs) (Fang et al. 2009). Similar effects may be detected in loci that are linked to sites under artificial selection, as selective sweeps can drive fixation of neutral or nearly neutral linked alleles (Innan and Kim 2004; Kim and Nielsen 2004; Rubin et al. 2010). Artificial selection could also

have genome-wide impacts on the rates and patterns of molecular evolution if selection for novelty promotes the evolution of mechanisms that increase the production of variation. For example, Burt and Bell (1987) found that domesticated mammals have higher chiasmata frequencies than other mammals with similar ages of maturity, which they suggested reflects "adaptation to an environment characterised by intense selection in small populations for novel combinations of traits". Otto and Barton (2001) also found several examples across different kingdoms that suggest a link between artificial selection regimes and increased recombination. Strong directional selection pressure and/or reduced effective population size could potentially increase the mutation rate (Sniegowski et al. 1997; Lynch 2010, 2011), though any increase in production of novel traits comes at the cost of a higher rate of deleterious mutations (King and Kashi 2007). While mitochondrial genomes of mammals and birds rarely if ever recombine, if domestication does indirectly select for generation of variation through recombination or mutation (Burt and Bell 1987; Denamur and Matic 2006; Dobney and Larson 2006; Bromham 2009), it could potentially influence rates of molecular evolution.

Relaxed selection could influence molecular evolution in domesticated lineages by permitting a greater proportion of non-synonymous mutations to persist. Some of the traits that experience relaxed selection during domestication may be related to changes in environmental conditions and lifestyle (Clutton-Brock 1999; Björnerfeldt et al. 2006; Driscoll et al. 2009; Rubin et al. 2010). For example, the higher proportion of non-synonymous changes in the mitochondrial genomes of dogs (Björnerfeldt et al. 2006) and domestic yaks (Wang et al. 2011) has been attributed to relaxed selection on metabolic efficiency in domesticated lineages, due to humans changing their habitat, selecting for tameness, and providing protection from predators.

Domesticated populations may often experience reductions in effective population sizes due to inbreeding and genetic bottlenecks (Vilà et al. 2005; Xia et al. 2009). Reduced effective population size increases the chance of fixing slightly deleterious mutations through drift, which should be reflected in increased d_N/d_S (Kimura and Ohta 1971; Ohta 1992). This effect is thought to account for patterns such as the correlation between body size and d_N/d_S in mammals (Nikolaev et al. 2007; Popadin et al. 2007; Nabholz et al. 2013). Domesticated lineages may undergo extreme bottlenecks on foundation. However, the domestication process has likely occurred over long periods of time, and may have included few or many bottlenecks interspersed with introgression

and population expansion (Allaby et al. 2008; Meyer and Purugganan 2013). This process could allow a lineage to recover from dramatic bottlenecks (Vilà et al. 2005). For example, although Taurine cattle may have originally descended from less than one hundred female founders (Bollongino et al. 2012), the high level of current genetic diversity has led to estimates of an ancestral wild population of 90,000 (McEachern et al. 2009). Ongoing selective breeding and narrowing of the breeding pool may have also reduced effective population size in some domesticated lineages (Medugorac et al. 2009). For example, dogs are likely to have experienced a prehistoric bottleneck from wolves (Vilà et al. 1997), but it is likely that some dog populations have experienced more severe bottlenecks in recent history from breeding pressure (Wayne and Ostrander 2007).

Changes in population structure or conditions during domestication may be expected to have significant impacts on molecular evolution. However, the generality of the relationship between domestication and patterns of molecular evolution has not been established. Is it confined to a few well-studied examples, or is it a more general feature of all domesticated lineages? Not all studies support higher non-synonymous rates in domestic lineages. For example, Rokas (2009) found a lower d_N/d_S in the proteome of a domesticated fungus compared to its wild relative. Here, we aim to ask whether increased d_N/d_S is a general feature of the mitochondrial genomes of domesticated lineages by comparing sequences from the maximum available number of phylogenetically independent comparisons of domesticated mammals and birds and their wild relatives.

We focus on the mitochondrial genome for several reasons. The animal mitochondrial genome has a higher rate of molecular evolution than the nuclear genome (Rand 1994; Ballard and Whitlock 2004), so is more likely to reflect any recent changes in rates and patterns of molecular evolution than the nuclear genome. The mitochondrial genome also has a smaller effective population size than the nuclear genome because it is haploid, rarely if ever recombines, and is maternally inherited (Harrison 1989; Moore 1995; Rokas et al. 2003; Ballard and Whitlock 2004), so it is expected to have a higher rate of fixation of nearly neutral substitutions (Ohta 1992), which are thought to dominate mitochondrial genome evolution (Rand and Kann 1996; Bazin et al. 2006). Since our aim is to include as many independent domestic lineages as possible, there is

a much wider availability of mitochondrial genomes than whole nuclear genome sequences.

To test whether domesticated animals have significantly different patterns of molecular evolution in mitochondrial genomes, we compared complete or nearly complete mitochondrial genome sequences between 16 phylogenetically independent comparisons of domesticated mammals and birds and their close wild relatives. We took two complementary approaches to analysing the data. We used a sister pairs approach to compare branch length, synonymous and non-synonymous differences and their ratios in wild and domesticated lineages. We also analysed all taxa together in a single phylogenetic ("whole tree") analysis. We found no evidence of a consistent difference between rates and patterns of molecular evolution in the mitochondrial genomes of domesticated mammals and birds and their wild relatives.

Methods

Selection of comparisons

We defined domesticated lineages as genetically distinct populations of organisms that have been purposely bred to suit the needs of the domesticator (Blumler et al. 1991; Diamond 2002). We identified the wild relatives of each domesticate from the literature, and collected information on the age and history of each domestication event (see Supplementary Material). We verified using published sources that the chosen wild relative and domestic populations could be identified as well supported, independent lineages from genetic data and that the domesticated and wild taxa were considered distinct based on morphology, behaviour, or geography.

To maintain phylogenetic independence among comparisons of domesticates and their wild relatives, we did not include multiple domesticated lineages that share the same wild relatives. For example, the llama and alpaca are suspected of sharing a wild relative (Kadwell et al. 2001; Cui et al. 2007), so we could only use one of these domesticates in our study. However, we were able to obtain whole mitochondrial genomes associated with two independently domesticated lineages for the dog

(Björnerfeldt et al. 2006) and the pig (Wu et al. 2007). For both the dog and pig, the two domesticate-wild comparisons were analysed as quartets, where one domesticate-wild relative pair acted as the outgroup for the other comparison.

DNA sequences

We found 16 comparisons of domesticates and their wild relatives with complete or nearly complete mitochondrial genome sequences available on GenBank (www.ncbi.nlm.nih.gov/genbank/). For each comparison, we collected a complete or nearly complete mitochondrial genome sequence for the domesticate, its wild relative, and a closely related outgroup (see Supplementary Table 3 for accession numbers and alignment lengths). We preferentially collected sequences for the most closely related wild relative and outgroup for each domesticated lineage for which we could obtain a complete or nearly complete mitochondrial genome sequence. We preferentially selected sequences from published papers that explicitly stated if the sequences came from wild or domesticated individuals. Sequences were not always available for the closest known wild relative, so in some cases we had to choose a more distant wild taxon. We conducted analyses with and without these more distant comparisons (for details see Supplementary Material). Similarly, in some cases there is evidence in the literature for population bottlenecks in the wild relatives, and this parallel change may make it harder to detect any effect of reduction in effective population size in the domesticated lineages. We repeated the sister pair and whole tree analyses excluding these comparisons to account for these potentially problematic comparisons.

We used a single mitochondrial genome to represent each taxon. This is because we wished to maximize the number of independent comparisons included in order to gauge general patterns of mitochondrial evolution in domesticated lineages. Multiple sequences are available for relatively few appropriate comparisons, and in many cases the lineages are not clearly monophyletic, which complicates the comparison of rates of substitution or levels of polymorphism (Hughes 2013). Use of a single sequence also avoids the problem of node density effect (Hugall and Lee 2007), especially because the level of polymorphism or number of substitutions may be overestimated in domesticated lineages if a greater number of sequences from domesticated lineages are included than sequences from the wild relatives. By using only a single sequence per

lineage, we are unable to distinguish between substitutions (present in all members of a population) and polymorphisms (present in some but not all members of a population).

Sister pairs analysis

We aligned the mitochondrial sequences (including protein-coding genes, rRNA, tRNA, and control region sequences) for each domesticate-wild relative comparison and outgroup. We also constructed alignments of only protein-coding genes for estimating non-synonymous (d_N) and synonymous (d_S) substitution rates. All alignments were performed by eye in Geneious (Drummond et al. 2011). We deleted any sites or codons that contained gaps in either the domesticate or wild relative sequence so that each base was comparable between sister species, and thus informative for a sister pairs analysis.

For the whole genome alignments for each comparison, we estimated branch lengths in BASEML (Yang 2007) using the TN93 substitution model and unconstrained rates (clock = 0 in PAML). We estimated d_S, d_N, and d_N/d_S for the protein-coding sequences in CODEML in PAML (version 4.4b, Yang 2007), using the F3x4 codon frequency model (clock = 0). We tested for significant differences in branch length for each comparisons using a likelihood ratio test (LRT).

We combined all 16 independent comparisons into a single analysis in order to ask whether the domesticated lineages have consistently different patterns of molecular evolution than their wild relatives. Each independent comparison contributed one data point to a non-parametric analysis of the differences in branch length, d_N, d_S and d_N/d_S between domesticates and their wild relatives. We used both a sign test and the Wilcoxon signed-ranks test (Wilcoxon 1946).

Since older divergences have had more time to accumulate substitutions, it may be that the power to detect a significant difference increases over time. If this were the case, we expect that if we compare age of domestication (years before present) or divergence of each sister pair (sum of domesticate and wild relative branch lengths) with the difference between domesticate and wild relative in d_N/d_S, d_N, d_S and total substitution rate, we would find that the older or more divergent comparisons are more likely to show a positive association between domestication and molecular evolution. To test this prediction, we used Spearman's rank correlation to test for an association between mean

age of domestication (measured in years before present, Table 1) and differences in branch lengths d_S, d_N, and d_N/d_S between domesticates and wild relatives. We also used Spearman's rank correlation to test for an association between the genetic distance between domesticated and wild lineages (measured as the sum of both the domestic and wild branches in each comparison) and differences in branch length, d_S, d_N, and d_N/d_S.

Whole tree analysis

In addition to the sister pairs approach, we performed a whole tree analysis where we combined the domesticated and wild taxa together into a single phylogeny. Because not all sequences could be confidently aligned between birds and mammals, we created three different alignments: (1) all sequences for all bird taxa; (2) all sequences for all mammal taxa; and (3) protein-coding sequences for all birds and mammals. The D-loop region was excluded from the whole tree analysis because it could not be confidently aligned across all taxa and was not available for several of the domesticate-wild relative comparisons.

For each of these three alignments, we estimated a phylogeny using the following procedure. First, we established data partitions for each alignments using a greedy search in PartitionFinder v1.0.1 (Lanfear et al. 2012), with linked branch lengths, constraining the models of evolution to those available in RAxML, and using AICc for model selection (a measure of AIC corrected for small sample sizes, Hurvich and Tsai 1989). In PartitionFinder we defined initial data blocks that separated protein-coding genes by gene and codon position. For alignments 1 (all bird genes) and 2 (all mammal genes), we treated the 12S and 16S rRNA genes as separate data blocks, and combined all tRNA sequences into one data block. Then, using the best partitions identified with PartionFinder, we analysed the three alignments in RAxML version 7.0.4 (Stamatakis et al. 2008) to estimate a maximum likelihood phylogeny for each alignment, with 1000 bootstrap replicates generated using the rapid bootstrapping algorithm. For the phylogenies based on alignments 1 and 2, we estimated branch lengths in BASEML (Yang 2007) using the REV model, unpartitioned data, and no molecular clock (clock = 0). For the phylogeny based on the protein-coding genes for birds and mammals, we used CODEML (Yang 2007) to estimate d_N/d_S in domesticated and wild lineages using the F3x4 codon frequency model, unpartitioned data and no molecular clock (clock = 0).

For all phylogenies, we then tested for a significant difference in branch length between domesticated lineages and non-domesticated lineages using a likelihood ratio test (LRT), comparing a one-rate model, where all taxa have the same rate, and a two-rate model, where one rate was estimated for all domesticates and a second rate for all wild relatives. A significant result from the LRT would allow us to reject the hypothesis of uniform rates over the phylogeny.

All alignment and data files used in this analysis are available on Dryad (http://datadryad.org) and can also be obtained from the corresponding author.

Results

Sister pairs analysis

We analysed differences in branch length, synonymous (d_S) and non-synonymous (d_N) differences, and d_N/d_S for 16 sister pairs between domesticated birds and mammals and their wild relatives using a sign test and Wilcoxon signed-rank test (Table 1). We found no evidence for a consistent difference between domesticated and wild lineages in branch length (sign test $p = 0.80$, Wilcoxon signed-ranks $p = 0.32$), synonymous rates (d_S: sign test $p = 0.46$, Wilcoxon signed-ranks $p = 0.78$), non-synonymous rates (d_N: sign test $p = 1.00$, Wilcoxon signed-ranks $p = 1.00$), nor d_N/d_S (sign test $p = 1.00$, Wilcoxon signed-ranks $p = 0.75$).

Six out of 16 comparisons showed a significant difference in branch length between the domesticated and wild lineages (presented in bold in Table 1). In three of these comparisons (llama and both pig lineages), the domesticated lineages had a significantly longer branch length. In the remaining three comparisons (sheep, cow goose), the wild relative had a significantly longer branch length.

A Spearman's rank correlation test revealed no evidence of a correlation between the age of the domestication event and direction of the difference between domesticated and wild lineages relatives in branch length ($\rho = 0.01$, $p = 0.97$), d_S ($\rho = 0.08$, $p = 0.76$) d_N (ρ

= 0.40, p = 0.13) nor in d_N/d_S (ρ = 0.14, p = 0.62). We also found no evidence of a correlation between domestication age and genetic distance between sister pairs (ρ = 0.01, p = 0.96), suggesting that, in the mitochondrial genome, the older comparisons included in this study do not always have the greatest genetic distance.

We found no significant relationship between genetic distance (sum of wild and domesticate branch lengths) and difference in d_S (ρ = -0.26, p = 0.34), d_N (ρ = 0.07, p = 0.80) or d_N/d_S
(ρ = 0.07, p = 0.79), but we did find a significant negative relationship between genetic distance and difference in branch length (ρ = -0.65, p = 0.01). This suggests that in the most divergent comparisons, the wild relative is more likely to have the longer branch length. The relationship is robust to the removal of either the cat or the cow comparisons, which are the most divergent comparisons (Figure S1); however, removing both of these comparisons makes the relationship non-significant. This result suggests that the net amount of molecular change between the sequences could influence the chance of detecting a difference in rate between the domesticated and wild relatives, but that this effect is unlikely to be responsible for our failure to detect more genetic change in domesticated lineages, since the relationship is in the opposite direction (greater genetic distance is associated with longer branches in the wild relative).

Whole tree analysis

For the whole tree analysis, we found no significant difference between the one and two-rate models for any of the three alignments we tested: (1) no significant difference in d_N/d_S for the alignment of protein-coding genes for all birds and mammals (p = 0.42); (2) no significant difference in branch length for whole genome alignment for all birds (p = 0.98); (3) no significant difference in branch length for whole genome alignment for all mammals (p = 0.95).

We repeated the sister pair and whole tree analyses, removing comparisons for which we were unable to use the closest wild relatives (cat, goat, cow, water buffalo and goose; either because of sequence availability or because the closest relative is extinct), or where we found evidence in the literature that the wild relatives have experienced genetic bottlenecks (camel, pigs, horse, donkey and water buffalo, see Supplementary

Material and Tables S1 and S2). Only a small part of the divergence in the distant comparisons (cat, goat, cow, water buffalo and goose) may actually correspond to molecular changes influenced by domestication, which could make these comparisons less informative. Furthermore, if reduced effective population size influences molecular rates in domesticates, we may have had difficulty detecting that signal when comparing a domesticate with a wild relative that has also experienced reduced effective population size. We repeated the analyses removing comparisons with suspected bottlenecks in the wild relatives: camel, pigs, horse, donkey and the water buffalo. We repeated this analysis with and without the water buffalo since the wild relative, the lowland anoa, has only recently experienced a genetic bottleneck (see Supplementary Material). In addition to experiencing a recent genetic bottleneck, the lowland anoa is an island endemic, which could be associated with a reduced effective population size and, thus, increased molecular rates (Woolfit and Bromham 2005).

When repeating the sister pairs analysis without comparisons where the wild relative has experienced bottlenecks (camel, pigs, horse, donkey and the water buffalo), we found that domesticates have a significantly higher d_N/d_S than their wild relatives (Wilcoxon signed-ranks, $p = 0.02$). Therefore, it is possible that in the pairs with bottlenecks in the wild relatives, reduced effective population size has had parallel effects in both domesticated lineages and their wild relatives, reducing the chance of detecting differences between them. All other alternative sister pair and whole tree analyses were not significant (Tables S1 and S2).

Discussion

We find no evidence for a general and consistent difference in the tempo and mode of mitochondrial molecular evolution of domesticated birds and mammals when compared to their wild relatives. Given that higher d_N/d_S has been reported for a number of domestic lineages, why do we fail to find evidence for a general increase in d_N/d_S across all the domestic lineages included in this study?

It is possible that lack of statistical power has prevented us from identifying significant differences in some comparisons. Our power is unavoidably limited by the nature of the

question. We are unable to include more comparisons because there are relatively few fully domesticated animal lineages, and we had to leave some lineages out of this study due to lack of sequence data from appropriate wild relatives (e.g., turkey: see Supplementary Material). It may be informative to apply this comparative approach to domesticated plants, which are more diverse. Furthermore, all domestication events are young on an evolutionary scale, so there has been only a short period of time for differences in tempo and mode of molecular evolution to make a detectable impression on patterns of sequence differences.

If the relatively small number of sequence differences between recently diverged genomes was obscuring a result, then we would expect the six sister pair comparisons with a significant difference in branch length to be more likely to show longer branch lengths, or higher d_N/d_S, in the domesticated lineage. But only half of the comparisons with a significant difference in branch length show more genetic change in the domesticated lineage, a pattern indistinguishable from chance. We also find that in the more divergent comparisons (those with a greater net genetic distance between the domestic and wild lineages), it is the wild relative that is more likely to have a longer branch length. Older domesticated lineages, that have had more time to accumulate evidence of distinct patterns of molecular evolution, do not show a greater tendency to have higher rates of change than their wild relatives. So we do not think that lack of power to detect differences in rate of change explains the lack of a consistent pattern in our comparisons. However, it may be possible that bottlenecks in wild relative populations may impact our power to detect a difference in d_N/d_S between domesticated lineages and their wild relatives.

One way to increase power to detect changes in the tempo and mode of molecular evolution in domesticated lineages is to take a population-level approach, with multiple individual samples for each domesticate and wild lineage. Recent population-level studies have found increased d_N/d_S or ratio of non-synonymous to synonymous diversity (π_N/π_S) in a number of domesticated lineages compared to their wild relatives (Wang et al. 2011; Hughes 2013). However, these studies have included an uneven number of domesticated and wild samples (254 from the dog vs. 19 from the wolf; 59 from the domestic pig vs. 27 from wild boar; 41 from the domestic chicken vs. 17 from the red junglefowl in Hughes 2013, and 51 domestic yaks vs. 21 wild yaks in Wang et al. 2011). Many short, recently-diverged branches can increase estimates of d_N/d_S (Rocha

2006), so higher d_N/d_S is more likely to be reported if an analysis includes more branches in a domesticate population than a wild one.

To avoid the measurement bias due to the node density effect, we only sampled one individual per domesticated and wild lineage. Choosing only one sequence per lineage also helps us to avoid the problem of lack of monophyly in analyses of population level data. Backcrossing and interbreeding with wild relatives can shape the molecular evolution of domesticated and wild lineages (Vilà et al. 2005), and these processes may have varied substantially between lineages. For example, Hughes (2013) reported that the phylogenies of domesticated and wild lineages chickens, dogs and pigs are not monophyletic but intermixed, which could be a signature of ancestral polymorphisms or interbreeding in these populations. We have attempted to minimise this effect on our results by choosing wild lineages that may not be the closest relative, but have less chance of being influenced by recent introgression (see Methods and Supplementary Material). However, by choosing only one sequence per lineage, we are unable to distinguish substitutions from polymorphisms. Our approach could mask higher rates of change in the domesticated lineage if wild lineages consistently retained comparatively more ancestral polymorphisms.

If the majority of substitutions in the mitochondrial genome are neutral or slightly deleterious, rather than under positive selection, then we would expect d_N/d_S estimates in the mitochondrial genome to be higher within species than between species (Hasegawa et al. 1998; Rand and Kann 1998; Weinreich and Rand 2000; Ho et al. 2005). Therefore, population-level estimates of mitochondrial d_N/d_S that do not account for the effect of ancestral polymorphism are expected to be higher than those estimated at the lineage-level. As such, we would expect our d_N/d_S estimates to be lower than those from population-level studies. Concordant with these population-level studies, we found a higher d_N/d_S in one dog, one pig, and the yak comparison. Although we can't compare our d_N/d_S estimates to the π_N/π_S reported in Hughes (2013), our d_N/d_S estimate for the domesticated and wild yaks are, as expected, lower than those reported by Wang et al. (2011) (our d_N/d_S for wild yaks: 0.06, their d_N/d_S for wild yaks: 0.07, our d_N/d_S for domesticated yaks: 0.09, their d_N/d_S for domesticated yaks 0.23).

It could be argued that the housekeeping genes of the mitochondria are unlikely to experience a dramatic change in selective regime, which could explain why we found

no consistent pattern associated with domestication in the mitochondrial genome. Actually, many studies of domestication report changes in traits associated with metabolism (Xia et al. 2009; Gibbons et al. 2012). For example, selective sweeps in chickens raised for meat production are connected to genes associated with growth, appetite and metabolic regulation (Rubin et al. 2010). It is therefore possible that artificially selected traits could be associated with growth and metabolism, which could potentially increase d_N in mitochondrial loci (MacEachern et al. 2009; Rubin et al. 2010; Akey et al. 2010; Amaral et al. 2011; Kijas et al. 2012). However, our study is designed to detect changes in genome-wide rates of change, rather than focusing on the effect of selection on particular genes.

Our results do not preclude an impact of domestication on patterns of mitochondrial evolution, but they do suggest that there is no consistent, detectable difference between all domesticated lineages and their wild relatives. It may be that domestication can influence mitochondrial molecular evolution, but that it does not do so consistently and uniformly across all domesticated lineages in comparison to their wild relatives. Each domestication history has involved different levels of human intervention, and the observed genetic and morphological changes in domesticated lineages are variable (Zeder 2006). For example, it has been suggested that domestic sheep and cats may have undergone less severe genetic bottlenecks than other domesticated animals (Driscoll et al. 2007; Kijas et al. 2009), but since both of these lineages have higher d_N/d_S estimates (Table 1) this does not seem to provide an explanation for the lack of a general pattern of higher d_N/d_S across domesticated lineages. Similarly, some domesticated lineages, like the horse, cat and camel may have experienced less artificial selection than others (Clutton-Brock 1999; Driscoll et al. 2009), yet the horse and cat have higher d_N/d_S than their wild relatives, and the camel has lower d_N/d_S.

In addition to considering the heterogeneity of processes affecting the domesticated lineages, population processes in the wild relatives may also impact on our ability to detect changes in the tempo and mode of molecular evolution in domesticated lineages. If similar changes have occurred in the both the domesticated lineages and their wild relatives, then we may be unable to detect a significant difference between them. In particular, some wild relatives may have experienced significant genetic bottlenecks. For example, the wild relative of the water buffalo, the lowland anoa, is an island endemic, which could be associated with a reduced effective population size and, thus,

increased d_N/d_S (Woolfit and Bromham 2005). Other examples of wild relatives that may have undergone population size reduction are the wild Bactrian camels (Hare 1997; Silbermayr et al. 2010), wild boar (Scandura et al. 2008), and Przewalski's horses (Clutton-Brock 1999; Vilà 2001) and the Somali wild ass (Moehlman 2002). When we analysed a reduced set of comparisons, removing comparisons where we found evidence that the wild relative had undergone a population bottleneck, we found that domesticated lineages had a higher d_N/d_S than their wild relatives. Although the sample size for this test is small (N =10), this result is consistent with the hypothesis that domestication reduces a lineage's effective population size and thus may increase the accumulation of slightly deleterious, non-synonymous changes in the mitochondrial genome.

In this analysis of 16 domesticated mammals and birds, we find no evidence of a general, consistent pattern in the rates or patterns of molecular evolution in the mitochondria. However, we do find that in a subset of comparisons, there is evidence of higher d_N/d_S in domesticated lineages, which may be a signature of changes in effective population size. We conclude that differences in d_N/d_S between particular domesticated lineages and their wild relatives in the mitochondrial genome (Björnerfeldt et al. 2006; Wang et al. 2011) are best explained by specific factors in the biology or domestication history of particular lineages, and not a generally predictable result of domestication.

Table 1: Comparison of mitochondrial genomes from 16 domesticated mammals and birds and their wild relatives. The age of domestication was calculated from the mean of published estimated ranges of timing of domestication events in years before present (see Supplemental Material for sources and details on the domesticate-wild comparisons chosen). Estimates of synonymous (d_S) and non-synonymous (d_N) substitutions rates, d_N/d_S, and total substitution rate (substitutions per site) were estimated in PAML v 4.4b (Yang 2007). Comparisons with significantly different branch lengths between domesticated and wild lineages (see Methods) are presented in bold. Domesticate-wild relative comparisons represented by two independent lineages (the dog and the pig), are marked with superscript one and two. Columns marked with D represent estimates for domesticates, and columns labelled W represent estimates for wild relatives. The sign columns represent the sign of the difference between domesticate and wild relative values. Positive symbols represent values that are larger in domesticates compared to their wild relatives, and negative symbols represent smaller values in domesticates.

† While these values are nearly equal, there is a small difference reflected in the direction of the sign of the d_S (chicken 0.003297, red junglefowl 0.003293) and d_N (chicken 0.000357, red junglefowl 0.000362). When the chicken comparison is removed from the dataset, the sign and Wilcoxon signed-ranks tests are still not significant.

Domesticate (Common name)	Domesticate (Scientific name)	Wild relative (Scientific name)	Mean domestication age	d_N/d_S			d_S			d_N			Branch length		
				D	W	Sign	D	W	Sign	D	W	Sign	D	W	Sign
Dog	*Canis lupus familiaris* [1]	*Canis lupus* [1]	13500	0.1673	0.1006	+	0.0083	0.0082	+	0.0014	0.0008	+	0.0045	0.0047	−
Dog	*Canis lupus familiaris* [2]	*Canis lupus* [2]	13500	0	0	−	0.0038	0.0019	+	0	0	−	0.0016	0.0010	+
Cat	*Felis catus*	*Lynx rufus*	9750	0.0330	0.0278	+	0.2811	0.3106	−	0.0093	0.0086	+	0.0814	0.0905	−
Horse	*Equus caballus*	*Equus przewalskii*	6000	0.0441	0.3065	−	0.0057	0.0030	+	0.0002	0.0009	−	0.0021	0.0013	+
Donkey	*Equus asinus*	*Equus asinus somalicus*	5000	0.0465	0.0648	−	0.0325	0.0234	+	0.0015	0.0015	−	0.0098	0.0065	+
Pig	***Sus scrofa domesticus*** [1]	***Sus scrofa*** [1]	9000	0.1728	0.1932	−	0.0136	0.0096	+	0.0024	0.0019	+	**0.0056**	**0.0037**	+
Pig	***Sus scrofa domesticus*** [2]	***Sus scrofa*** [2]	9000	0.0575	0.0497	+	0.0129	0.0074	+	0.0007	0.0004	+	**0.0033**	**0.0018**	+
Goat	*Capra hircus*	*Capra falconeri*	11000	0.0662	0.0601	+	0.0474	0.0476	−	0.0031	0.0029	+	0.0141	0.0176	−
Sheep	***Ovis aries***	***Ovis ammon***	8500	0.0562	0.0391	+	0.0369	0.0596	−	0.0021	0.0023	−	**0.0135**	**0.0155**	−
Cow	***Bos taurus***	***Bos gaurus***	9000	0.0360	0.0400	−	0.1566	0.1486	+	0.0056	0.0059	−	**0.0405**	**0.0426**	−
Yak	*Bos grunniens*	*Bos grunniens*	4750	0.0870	0.0634	+	0.0090	0.0123	−	0.0008	0.0008	+	0.0034	0.0038	−
Water buffalo	*Bubalus bubalis*	*Bubalus depressicornis*	5000	0.0476	0.1009	−	0.0447	0.0429	+	0.0021	0.0043	−	0.0099	0.0162	−
Llama	***Lama glama***	***Lama guanicoe***	6000	0.0431	0.0171	+	0.0378	0.0313	+	0.0016	0.0005	+	**0.0113**	**0.0080**	+
Bactrian camel	*Camelus bactrianus*	*Camelus ferus*	5000	0.0591	0.0730	−	0.0305	0.0329	−	0.0018	0.0024	−	0.0089	0.0102	−
Chicken	*Gallus gallus*	*Gallus gallus gallus*	8250	0.1083	0.1098	−	0.0033†	0.0033†	+	0.0004†	0.0004†	−	0.0011	0.0010	+
Goose	***Anser anser***	***Anser albifrons***	3500	0.0511	0.0302	+	0.0310	0.0443	−	0.0016	0.0013	+	**0.0071**	**0.0122**	−
		No. positive comparisons				8			10			8			7
		Total no. comparisons				16			16			16			16
		Sign test *p*				1			0.45			1			0.80
		Wilcoxon signed-rank test *p*				0.75			0.78			1			0.32

References

Allaby RG, Fuller DQ, Brown TA. 2008. The genetic expectations of a protracted model for the origins of domesticated crops. Proc Natl Acad Sci U S A. 105:13982–13986.

Akey JM et al. 2010. Tracking footprints of artificial selection in the dog genome. Proc Natl Acad Sci U S A. 107:1160–1165.

Amaral AJ, et al. 2011. Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. PLoS ONE 6:e14782.

Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. Mol Ecol. 13:729–744.

Bazin E, Glémin S, Galtier N. 2006. Population size does not influence mitochondrial genetic diversity in animals. Science 312:570–572.

Bollongino R, et al. 2012. Modern Taurine cattle descended from small number of Near-Eastern founders. Mol Biol Evol. 29:2101–2104.

Björnerfeldt S, Webster MT, Vilà C. 2006. Relaxation of selective constraint on dog mitochondrial DNA following domestication. Genome Res. 16:990–994.

Blumler MA, et al. 1991. The ecological genetics of domestication and the origins of agriculture. Curr Anthropol. 32:23–54.

Bromham L. 2009. Why do species vary in their rate of molecular evolution? Biol Lett. 5:401–404.

Burt A, Bell G. 1987. Mammalian chiasma frequencies as a test of two theories of recombination. Nature 326:803–805.

Clutton-Brock J. 1999. *A natural history of domesticated mammals*. 2nd ed. Cambridge University Press, Cambridge, U.K.

Cruz F, Vilà C, Webster MT. 2008. The legacy of domestication: Accumulation of deleterious mutations in the dog genome. Mol Biol Evol. 25:2331–2336.

Cui P, et al. 2007. A complete mitochondrial genome sequence of the wild two-humped camel (*Camelus bactrianus ferus*): An evolutionary history of Camelidae. BMC Genomics 8:241.

Denamur E, Matic I. 2006. Evolution of mutation rates in bacteria. Mol Microbiol. 60:820–827.

Diamond J. 2002. Evolution, consequences and future of plant and animal domestication. Nature 418:700–707.

Dobney K, Larson G. 2006. Genetics and animal domestication: New windows on an elusive process. J Zool. 269:261–271.

Driscoll CA, Macdonald DW, O'Brien SJ. 2009. From wild animals to domestic pets, an evolutionary view of domestication. Proc Natl Acad Sci U S A. 106:9971–9978.

Driscoll CA, et al. 2007. The Near Eastern origin of cat domestication. Science 317:519–523.

Drummond A, et al. 2011. Geneious v5.4. Available from http://www.geneious.com/.

Fang M, Larson, G, Soares Ribeiro H, Li N, Andersson L. 2009. Contrasting mode of evolution at a coat color locus in wild and domestic pigs. PLoS Genet. 5:e1000341.

Gibbons JG, et al. 2012. The evolutionary imprint of domestication on genome variation and function of the filamentous fungus *Aspergillus oryzae*. Curr Biol. 22:1403–1409.

Gu Z, et al. 2005. Elevated evolutionary rates in the laboratory strain of *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A. 102:1092–1097.

Hare J. 1997. The wild Bactrian camel *Camelus bactrianus ferus* in China: The need for urgent action. Oryx 31:45–48.

Harrison RG. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends Ecol Evol. 4:6–11.

Hasegawa M, Cao Y, Yang Z. 1998. Preponderance of slightly deleterious polymorphism in mitochondrial DNA: nonsynonymous/synonymous rate ratio is much higher within species than between species. Mol Biol Evol. 15:1499–1505.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol. 22:1561–1568.

Hugall AF, Lee MSY. 2007. The likelihood node density effect and consequences for evolutionary studies of molecular rates. Evolution 61:2293–2307.

Hughes AL. 2013. Accumulation of slightly deleterious mutations in the mitochondrial genome: A hallmark of animal domestication. Gene 515:28-33.

Hurvich CM, Tsai CL. 1989. Regression and time series model selection in small samples. Biometrika 76:297–307.

Innan H, Kim Y. 2004. Pattern of polymorphism after strong artificial selection in a domestication event. Proc Natl Acad Sci U S A. 101:10667–10672.

Kadwell M, et al. 2001. Genetic analysis reveals the wild ancestors of the llama and the alpaca. Proc. Biol. Sci. 268:2575–2584.

Kijas JW, et al. 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS ONE 4:e4668.

Kijas JW, et al. 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. 10:e1001258.

Kim Y, Nielsen R. 2004. Linkage disequilibrium as a signature of selective sweeps. Genetics 167:1513–1524.

Kimura M, Ohta T. 1971. *Theoretical aspects of population genetics*. Princeton University Press, Princeton, USA.

King DG, Kashi Y. 2007. Indirect selection for mutability. Heredity 99:123–124.

Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol. 29:1695–1701.

Lu J, et al. 2006. The accumulation of deleterious mutations in rice genomes: A hypothesis on the cost of domestication. Trends Genet. 22:126–131.

Lynch M. 2010. Evolution of the mutation rate. Trends Genet. 26:345–352.

Lynch M. 2011. The lower bound to the evolution of mutation rates. Genome Biol Evol. 3:1107–1118.

MacEachern S, et al. 2009. Molecular evolution of the Bovini tribe (Bovidae, Bovinae): Is there evidence of rapid evolution or reduced selective constraint in domestic cattle? BMC Genomics 10:179.

Medugorac I, et al. 2009. Genetic diversity of European cattle breeds highlights the conservation value of traditional unselected breeds. Molec Ecol. 18:3394–3410.

Meyer RS, Purugganan MD. 2013. Evolution of crop species: genetics of domestication and diversification. Nat Rev Genet. 14:840–852.

Moehlman, PD. 2002. Status and action plan for the African wild ass. In: Moehlman P. D., editor. *Equids: zebras, asses and horses, status survey and conservation action plan*. Gland, Switzerland: IUCN. p. 2–9.

Moore WS. 1995. Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees. Evolution 718–726.

Nabholz B, Uwimana N, Lartillot N. 2013. Reconstructing the phylogenetic history of long-term effective population size and life-history traits using patterns of amino acid replacement in mitochondrial genomes of mammals and birds. Genome Biol Evol. 5:1273–1290.

Nikolaev SI, et al. 2007. Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements. Proc Natl Acad Sci U S A. 104:20443–20448.

Nishibori M, Shimogiri T, Hayashi T, Yasue T. 2005. Molecular evidence for hybridization of species in the genus *Gallus* except for *Gallus varius*. Anim Genet. 36:367-375.

Ohta T. 1992. The nearly neutral theory of molecular evolution. Annu Rev Ecol Syst. 23:263–286.

Otto SP, Barton NH. 2001. Selection for recombination in small populations. Evolution 55:1921–1931.

Popadin K, Polishchuk LV, Mamirova L, Knorre D, Gunbin K. 2007. Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. Proc Natl Acad Sci U S A. 104:13390–13395.

Rand DM. 1994. Thermal habit, metabolic rate and the evolution of mitochondrial

DNA. Trends Ecol Evol. 9:125–131.

Rand DM, Kann LM. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. Mol Biol Evol. 13:735–748.

Rand DM, Kann LM. 1998. Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA. Genetica 102-103:393–407.

Rocha EPC, et al. 2006. Comparisons of d_N/d_S are time dependent for closely related bacterial genomes. J Theor Biol. 239:226–235.

Rokas A. 2009. The effect of domestication on the fungal proteome. Trends Genet. 25:60–63.

Rokas A, Ladoukakis E, Zouros E. 2003. Animal mitochondrial DNA recombination revisited. Trends Ecol Evol. 18:411–417.

Rubin CJ, et al. 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464:587–591.

Scandura M, et al. 2008. Ancient vs. recent processes as factors shaping the genetic variation of the European wild boar: Are the effects of the last glaciation still detectable? Mol Ecol. 17:1745–1762.

Silbermayr K, et al. 2010. High mitochondrial differentiation levels between wild and domestic Bactrian camels: A basis for rapid detection of maternal hybridization. Anim Genet. 41:315–318.

Sniegowski PD, Gerrish PJ, Lenski RE. 1997. Evolution of high mutation rates in experimental populations of *E. coli*. Nature 387: 703–705

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol. 57:758.

Vilà C. 2001. Widespread origins of domestic horse lineages. Science 291:474–477.

Vilà C, et al. 1997. Multiple and ancient origins of the domestic dog. Science 276:1687–1689.

Vilà C, Seddon J, Ellegren H. 2005. Genes of domestic mammals augmented by backcrossing with wild ancestors. Trends Genet. 21:214–218.

Wang Z, et al. 2011. Domestication relaxed selective constraints on the yak mitochondrial genome. Mol Biol Evol. 28:1553–1556.

Wayne RK, Ostrander EA. 2007. Lessons learned from the dog genome. Trends Genet. 23:557–567.

Weinreich DM, Rand DM. 2000. Contrasting patterns of nonneutral evolution in proteins encoded in nuclear and mitochondrial genomes. Genetics 156:385–399.

Wilcoxon F. 1946. Individual comparisons of grouped data by ranking methods. J Econ Entomol. 39:269.

Woolfit M, Bromham L. 2005. Population size and molecular evolution on islands. Proc R Soc Lond B Biol Sci. 272:2277-2282.

Wu GS, et al. 2007. Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multiple domestication events in East Asia. Genome Biol. 8:R245.

Xia Q, et al. 2009. Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). Science 326:433–436.

Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol. 24:1586–1591.

Zeder MA. 2006. Central questions in the domestication of plants and animals. Evol Anthropol. 15:105–117.

Supplementary Information

Here we provide information on the domesticated lineages and their wild relatives analysed in this study. For GenBank accession numbers used in this study, please see Supplementary Table 3 (http://gbe.oxfordjournals.org/content/6/1/161/suppl/DC1).

Dog (*Canis lupus familiaris*):

Dog domestication probably began 12000-15000 years ago and involved multiple spatially and temporally separate domestication events (Clutton-Brock 1999, Vilà et al. 1997; Savolainen et al. 2002). However, a recently reported skull from the Altai Mountains has led to claims of a much earlier origin ca. 33,000 years ago (Druzhkova et al. 2013). A phylogeny produced by Björnerfeldt et al. (2006) using whole mitochondrial genomes supports multiple domestication of dogs (*Canis lupus familiaris*) from grey wolves (*Canis lupus*). In two of the four dog-wolf clades identified, the dog and wolf branches are clearly resolved and well supported as sister species: one in clade II and another in clade IV. We used whole mitochondrial genome sequences published by Björnerfeldt et al. (2006) for two independent dog comparisons, selecting one dog-wolf comparison randomly from clade II (D4, W4) and one from clade IV (D12, W3). Because each clade represents a separate origin of a domesticated lineage, we treat the two dog-wolf comparisons as separate data points in our analysis. We analysed these comparisons as a quartet, such that each pair acted as the outgroup to the other.

Cat (*Felis catus*):

The cat (*Felis catus*) was originally domesticated from the Near Eastern wild cat (*Felis silvestris libyca*) 9500-10000 years ago (Vigne et al. 2004; Driscoll et al. 2007). While many other domesticated lineages may have been chosen based on their tameability and use for meat or labour, the cat may have instead chosen to exploit humans for food scraps and house mice living around settlements (Clutton-Brock 1999; Driscoll et al. 2009).

There are no whole mitochondrial genome sequences available for any of the wild *Felis* members. We chose the *Lynx rufus* (bobcat) as the wild relative based on relatedness (Agnarsson et al. 2010) and sequence availability. We chose the leopard cat

(*Prionailurus bengalensis*) as the outgroup taxon based on the Agnarsson et al. (2010) phylogeny and sequence availability.

Horse (*Equus caballus*):

The horse (*Equus caballus*) was probably first domesticated near the Eurasian Steppe between 5000-7000 years ago, possibly first for meat and later for riding (Anthony and Brown 1991; Levine 1999). There are only a few wild horse types in historic records: the tarpans (*Equus ferus ferus*) and Exmoor ponies (*Equus ferus caballus*) which are believed to be feralized or hybrids with previously domesticated horses (Clutton-Brock 1999), and Przewalski's horse (*Equus przewalskii*) which is believed to be the closest wild, extant sister taxon of the domestic horse (Clutton-Brock 1999; Kavar and Dovč 2008; Goto et al. 2011). Przewalski's horse is now endangered and experienced a severe bottleneck when the population nearly went extinct in the wild (Ryder 1993).

The domestication history of the horse may have been characterised by many matrilines, high migration, no tight bottleneck, and less selective breeding than other domesticates (Vilà et al. 2001; Wade et al. 2009). Selective breeding may have had much less of an impact on the domestication of the horse, and domestication may have been driven by the spread of the methodology to capture and tame wild horses, which were becoming scarce as many as 10000 years ago (Anthony and Brown 1991; Clutton-Brock 1999; Levine 1999; Vilà 2001).

There are several wild populations of Przewalski's horse, and genetic evidence suggests that some have experienced historic hybridization with domestic horses while others are genetically distinct from domestic horses (Goto et al. 2011). Przewalski's horses are also morphologically distinct (Sasaki et al. 1999), being more similar to zebras and wild asses, and having a higher chromosome number than domestic horses (Benirschke et al. 1965; Sasaki et al. 1999; Wallner et al. 2003). Goto et al. (2011) suggest that the Bonnette variety of Przewalski's horse in particular is sister to the domestic horse, so for our analysis we used a complete mitochondrial sequence from a Bonnette variety Przewalski's horse (Haplotype III: AP012269) presented by Goto et al. (2011), which is less likely to contain a signal of interbreeding with the domestic horse. We chose a domestic horse lineage chosen at random from the same study (Thoroughbred, HQ439462). We used the onager (*Equus hemionus*) as the outgroup taxon.

Donkey (*Equus asinus*):

Donkey (*Equus asinus*) domestication most likely occurred in North Africa about 5000 years ago (Clutton-Brock 1999; Beja-Pereira et al. 2004). The African wild ass, which is split into the Nubian (*Equus asinus africanus*) and Somali (*Equus asinus somalicus*) subspecies, is believed to be the domestic donkey's closest living wild relative. Both subspecies are considered critically endangered due to overhunting and habitat and population fragmentation (Moehlman 2002). Genetic, morphological and geographic evidence suggest that the Nubian subspecies may be the closest wild relative to the domestic donkeys (Beja-Pereira et al. 2004; Rossel et al. 2008); however, other genetic studies reveal links between extant domesticated lineages and both Nubian and Somali populations (Aranguren-Mendez et al. 2004; Kimura et al. 2010). Studies also support that the Somali wild asses form a geographically and genetically distinct population (Aranguren-Mendez et al. 2004; Beja-Pereira et al. 2004). For our analysis, we used a nearly complete mitochondrial DNA sequence for the domestic donkey from Xu et al. (1996) and for the Somali wild ass published by Goto et al. (2011). There is no mitochondrial genome sequence for the Nubian wild ass on GenBank. We used Przewalski's horse (*Equus przewalskii*) as the outgroup taxon.

Pig (*Sus scrofa domesticus*):

The pig (*Sus scrofa domesticus*) was domesticated about 9000 years ago from the Eurasian wild boar (*Sus scrofa*), with a subsequent domestication event in Europe (Clutton-Brock 1999). According to Ramírez et al. (2009) and Amaral et al. (2011), there is little genetic difference between wild boars and domestic pigs, and domestic pigs are found to be more different from each other than they are to wild boars. This could have occurred through multiple domestications of the pig across Eurasia, allowing for the contribution of the wild pig genes into both Asian and European gene lines (Larson 2005). Wild boar populations may have suffered a series of bottleneck events due to overhunting (Scandura et al. 2008). Furthermore, there is a long history of interbreeding between wild boars and domesticated pigs (e.g., Ollivier 2009; Ramírez et al. 2009).

Genetic studies support independent European and Asian clades, and also independent lineages within Europe (Giuffra et al. 2000; Kijas and Andersson 2001). Wu et al. (2007) found a similar pattern using nearly complete mitochondrial genomes, and also revealed an independent East Asian domesticated lineage derived from Asian wild

boars. We have chosen one European and one Asian domestic-wild comparison from Wu et al. (2007) to represent two independent pig domestication events in our study. Like the dog, we analysed these comparisons as a quartet, using one pair as the outgroup to the other.

Goat (*Capra hircus*):

Of the multiple goat domestication events, the oldest occurred 10000-12000 years ago (Zeder 2000). Although there are several whole mitochondrial genome sequences available for the domesticated goat (*Capra hircus*), Hassanin et al. (2010) found errors in several of these sequences, so for this study, we used a newer mitochondrial genome sequence presented by Hassanin et al. (2010). The bezoar (*Capra aegagrus*) is the closest wild relative to the domesticated goat (Harris 1962; Zeder 2000), however, there are only a few genes available on GenBank. The next closest wild relative is the markhor (*Capra falconeri*) (Takada et al. 1997; Pidancier et al. 2006), which has been used in other molecular studies of the domestic goat and its relatives (Luikart et al. 2001; Hassanin et al. 2009; 2010) and is well supported as a sister lineage of the domestic goat (Hassanin et al. 2009; 2010). We used the whole mitochondrial genome sequence of *Capra falconeri* used by several studies (Luikart et al. 2001; Hassanin et al. 2009; 2009; 2010) for the wild relative. Using information from the same sources, we chose *Hemitragus jemlahicus* as the outgroup, using the sequence presented in Hassanin et al. (2010).

Sheep (*Ovis aries*):

Sheep (*Ovis aries*) have been domesticated multiple times, with the first event dating back 8000-9000 years ago (Clutton-Brock 1999). Researches have debated whether the closest wild relative of the domestic sheep is the European or Asiatic mouflon (*Ovis aries musimon* and *Ovis orientalis*), the argali (*Ovis ammon*), or the urial (*Ovis vignei*). Recent studies using mitochondrial DNA (Hiendleder et al. 1998; 2002; Pedrosa et al. 2005; Meadows et al. 2006), revealed five haplogroups in the domestic sheep. Meadows et al. (2010) published a phylogeny including all five haplogroups as well as samples from the European mouflon, the argali and the urial. Based on this phylogeny, the European mouflon is characterised as a feralized form of a previously domesticated lineage (Chessa et al. 2009), and the argali is considered the best candidate for the closest wild relative. For our study, we have chosen the argali as the wild relative of the sheep. We used sheep and argali sequences published by Meadows et al. (2010). We

selected a sequence from a domestic sheep carrying one of the most common haplogroups (HA), since this haplogroup is found in domestic sheep worldwide. We chose *Ammotragus lervia* (aoudad) for the outgroup (Hassanin et al. 2009).

Cow (*Bos taurus*):

Cattle were domesticated about 8000-10000 years ago (Loftus et al. 1994; Clutton-Brock 1999). Domestic cattle (*Bos indicus* and *Bos taurus*) are now generally considered to be subspecies of the extinct *Bos primigenius* (Bailey et al. 1996; Hiendleder et al. 2008). Since both domestic cattle lineages are likely to share the same wild relative, we only examined the Taurine cow (*Bos taurus*) in our analysis. Many of the close relatives of the cow in the genus *Bos* have been partially or fully domesticated (e.g., gayal, banteng). The gaur (*Bos gaurus*), however, is a good candidate for use as the wild relative of the cow in this study as it is closely related, has not been domesticated (Clutton-Brock 1999), and has a complete mitochondrial genome on GenBank. For the outgroup, both yak and bison are closely related to domestic cattle (e.g., Hassanin and Ropiquet 2004; MacEachern et al. 2009; Robinson and Ropiquet 2011), so we used the European bison or wisent (*Bison bonasus*). The European bison was likely the product of continuous interbreeding of an ancient cattle-like ancestor with other bison species (Verkaar 2004; Pertoldi et al. 2010), making it a closer relative of domestic cattle than the American bison. However, although in a different genera, we found that the sequences of *Bos gaurus* and *Bison bonasus* appeared to both have a similar degree of divergence from *Bos taurus* (a similar finding is presented in Robinson and Ropiquet 2011), so instead we used the wild water buffalo, *Bubalus bubalis*, for the outgroup of the cow and gaur.

Yak (*Bos grunniens*):

The yak (*Bos grunniens*) was most likely domesticated in Asia 4500-5000 years ago, though the date is still debated (Wiener et al. 2003; Rhode et al. 2007). Many publications note that the domestic and wild yaks are much more genetically similar to each other than other domestic and wild comparisons, which may be due to continual but low levels of interbreeding throughout the domestication history (Li et al. 2007; Wang et al. 2010b; 2011). The wild and domestic individuals, however, are easily distinguished morphologically and geographically. Wang et al. (2011) published a phylogeny including full mitochondrial sequences of 51 domestic and 21 wild geographically and morphologically distinct yaks. Their results reveal evidence of

possible multiple domestication events and interbreeding throughout the evolutionary history of the domestic yak. Three main clades are defined, one of a basal wild lineage, and two clades containing wild and domestic mixes. For our analysis, we used the most basal wild yak (W71) and a randomly chosen domestic yak from a clade containing only domestic individuals (HY1) from Wang et al. (2011). Because there is no strong evidence that particular mitochondrial genomes correspond to different domestication events (Guo et al. 2006; Wang et al. 2011), we only chose one domesticate-wild comparison for analysis in this study. We used the American bison (*Bison bison*) as the outgroup for this comparison.

Water buffalo (*Bubalus bubalis*):

The water buffalo (*Bubalus bubalis*) was first domesticated approximately 5000 years ago (Clutton-Brock 1999). Domestic water buffalo can be split into domestic river buffalo and domestic swamp buffalo. The two domestic water buffalo are able to interbreed, but have different karyotypes and are genetically distinct (Kikkawa et al. 1997; Kierstein et al. 2004; Yindee et al. 2010). The closest relative to both domesticates is the wild water buffalo (*Bubalus arnee*) (Clutton-Brock 1999), so we can only use one domesticated water buffalo for our study. There is no mitochondrial genome sequence available for *Bubalus arnee*, so instead we used the next closest wild relative, the lowland anoa (*Bubalus depressicornis*) (Hassanin and Ropiquet 2004). We used the African buffalo (*Syncerus cafer*) as the outgroup.

The lowland anoa is endemic to Sulawesi Island in Indonesia and has recently become endangered due to overhunting and habitat loss, particularly in the last 30 years (Burton et al. 2005). Because they are an island species, they may have a lower effective population size than their mainland relatives (e.g., wild water buffalo), which may lead to higher estimates of d_N/d_S (Woolfit and Bromham 2005). The recent bottleneck may also influence genetic diversity, though this effect may be too recent to detect. For these reasons, we repeat our analyses that exclude comparisons with wild relatives that have experienced genetic bottlenecks (Tables S1 and S2), with and without the water buffalo.

Llama (*Lama glama*):

The alpaca (*Vicugna pacos*) was domesticated 6000-7000 years ago in the Andes, most likely from the vicuña (*Vicugna vicugna*) (Wheeler 1995; Wheeler et al. 1995). The llama (*Lama glama*) was most likely domesticated from the guanaco (*Lama guanicoe*)

approximately 6000 years ago, however the history is debated (Hemmer 1990; Wheeler 1995).

Genetic evidence corroborates a recent history (ca. 25 years) of intentional hybridization between the llama and alpaca, with 80% of alpacas and 40% of llamas showing signs of hybridization (Kadwell et al. 2001). However, many studies support the genetic divide between llama and guanaco and the alpaca and vicuña (Kadwell et al. 2001; Marín et al. 2007a; 2007b). Although there is evidence for the genetic separation of the guanaco-llama and vicuña-alpaca lineages, there is a history of interbreeding between the llama and alpaca (Wheeler 1995). We chose to include only the llama, using the guanaco as the wild relative and the vicuña as the outgroup.

Bactrian Camel (*Camelus bactrianus*):

The Bactrian camel (*Camelus bactrianus*) was domesticated approximately 5000 years ago in Central Asia (Clutton-Brock 1999). The small population of wild Bactrian camels are very similar morphologically to their domestic counterparts, and historically, it was unclear whether they are a feralized population of escaped domesticates (Clutton-Brock 1999). Genetic evidence, however, suggests that *Camelus ferus* and the domestic *Camelus bactrianus* may represent distinct genetic lineages (Ji et al. 2009; Silbermayr et al. 2010). Evidence from complete mitochondrial genome sequences suggests that the wild and domestic Bactrian camels are genetically distinct and that *Camelus ferus* is a wild relative of *Camelus bactrianus* (Ji et al. 2009). Ji et al. (2009) also suggest that the level of genetic difference between the wild and domesticated Bactrian camels is evidence that they have different progenitors that diverged millions of years ago. *Camelus ferus* is critically endangered and has been reported as scarce or near extinction since the 1970s (Hare 1997). For this study, we used *Camelus ferus* as the wild relative of the domesticated Bactrian camel and the Dromedary camel (*Camelus dromedarius*) as the outgroup taxon.

The wild relative of the closely related Dromedary camel is unknown (Clutton-Brock 1999), making the Bactrian camel the next closest relative of the Dromedary camel. Thus, we only included the Bactrian camel in this analysis.

Chicken (*Gallus gallus*):

Based on archaeological evidence, the chicken (*Gallus gallus*) was most likely domesticated in Asia between 6500-10000 years ago (Shinan 1996). When contemplating the origin of the domestic chicken, Darwin (1868) noted that based on morphology, the domestic chicken most likely descended from the red junglefowl (*Gallus gallus*), as opposed to the green, grey, or ceylon junglefowl.

Genetic studies have provided evidence confirming the close relationship between the red junglefowl and the domestic chicken (e.g., Fumihito et al. 1994; 1996; Nishibori et al. 2005; Sawai et al. 2010), so we used the mitochondrial genome sequence from the red junglefowl (*Gallus gallus* subsp. *gallus*, RJF gal1,2) from Nishibori et al. (2005) as the wild relative in this study. Although *Gallus gallus* subsp. *bankiva* also appears to be closely related to the domestic chicken, we chose not to use this taxon since it originates from an island, meaning it may have a reduced effective population size compared to mainland junglefowl, which could influence rates of molecular evolution (Woolfit and Bromham 2005). This history could dampen our ability to compare a signal of genetic bottlenecks in domesticated chickens.

Nishibori et al. (2005) also identified genetic evidence of hybridization between the domestic chicken and all junglefowl except for the green junglefowl (*Gallus varius*), so in this study we used the green junglefowl for the outgroup.

Goose (*Anser anser*):

The goose was domesticated in Egypt 3000-4000 years ago (Buckland and Guy 2002; Wang et al. 2010a). It is likely that there are several origins of the domestic goose, at least one in Eurasia and two in Asia, each of which descended from different, but closely related, goose species (Shi et al. 2006; Zhu et al. 2010; Wang et al. 2010a). The greylag goose (*Anser anser*) was domesticated in Europe and the swan goose (*Anser cygnoides*) was domesticated in Asia (Buckland and Guy 2002; Wang et al. 2010a). Based on sequence availability, we chose the domestic greylag goose (*Anser anser*) for this study. There are no mitochondrial genome sequences available for wild greylag geese. The two domestic geese species and their wild relatives are closely related to *Anser albifrons*, *A. brachyrynchus*, *A. erythropis*, and *A. fabalis*, collectively forming the ‘grey geese’ (Ruokonen et al. 2000). Based on sequence availability, we could choose between *Anser albifrons* and *Anser fabalis* as the wild relative. Many sections of

the sequence for *Anser fabalis* were missing or could not be aligned as confidently across the domesticate and outgroup as the comparison with *Anser albifrons* (particularly sections of ATP6, COX2, ND1, ND6 and 16s). All results for the goose comparison presented in the manuscript are based on the comparison between the domesticated greylag goose and *Anser albifrons*. We chose the Canada goose, *Branta canadensis* as the outgroup based on relatedness (Donne-Gousse et al. 2002) and sequence availability.

Turkey (*Meleagris gallopavo*):

Because of poor data availability, this would have been a very distant comparison, so we chose not to include the turkey in this analysis. The turkey (*Meleagris gallopavo*) was domesticated 500-2200 years ago in North America, probably from at least two domestication events (Crawford 1992; Brant 1998; Speller et al. 2010). Although many subspecies of wild turkey may have contributed to modern domesticated lineages, the progenitors of the two domestication events are most likely the Eastern wild turkey (*Meleagris gallopavo silvestris*) and the South Mexican wild turkey (*Meleagris gallopavo gallopavo*) (Crawford 1992; Speller et al. 2010). Mitochondrial genome sequences are not available for any close relatives of the turkey. Based on recent phylogenies, the closest relative with a published mitochondrial genome is the hazel grouse (*Tetrastes bonasia*) (Kimball and Braun 2008; Shen et al. 2010; Zhao et al. 2012).

Table S1: Results of sister pairs analyses for subsets of the comparisons that exclude distant domesticate-wild relative comparisons or comparisons where the wild relatives have experienced a genetic bottleneck. The *p*-values from sign and Wilcoxon signed-ranks tests for the rate of non-synonymous substitutions (d_N), synonymous substitutions (d_S) and their ratio (d_N/ d_S) along with total substitution rate from all 16 comparisons reported in the manuscript are compared with results when 1) the cat is removed (the most distant pair), 2) all distant comparisons are removed (cat, goat, cow, water buffalo, goose), 3) all comparisons where the wild relative (WR) has experienced a bottleneck are removed (horse, camel, pig, donkey, water buffalo), and 4) when both distant comparisons and comparisons with WR bottlenecks are removed. Significant results ($p \leq 0.05$) are indicated with an asterisk. We found $d_N = 0$ for one dog pair (see Table 1), so we repeated the Wilcoxon signed-ranks tests without this data point (values in parentheses).

Subset	Sign test *p*-values				Wilcoxon signed-ranks test *p*-values			
	d_N/d_S	**d_S**	**d_N**	**Branch length**	**d_N/d_S**	**d_S**	**d_N**	**Branch length**
All comparisons	1	0.45	1	0.80	0.75 (0.76)	0.78	1 (0.80)	0.32
Cat removed	1	0.30	1	1	0.85 (0.85)	0.45	1 (1)	0.52
Cat and cow removed	1	0.42	1	1	0.83 (0.84)	0.63	0.89 (0.89)	0.71
Distant comparisons removed	1	0.23	1	0.55	0.84 (0.85)	0.32	0.84 (0.85)	0.28
Comparisons with WR bottleneck removed	0.55	1	1	0.23	0.13 (0.13)	0.64	0.61 (0.62)	0.07
Comparisons with WR bottleneck removed (water buffalo also removed)	0.34	1	1	0.34	0.02* (0.02) *	0.49	0.24 (0.25)	0.13
Distant and WR bottleneck comparisons removed	0.69	0.69	1	1	0.11 (0.12)	1	0.59 (0.62)	1

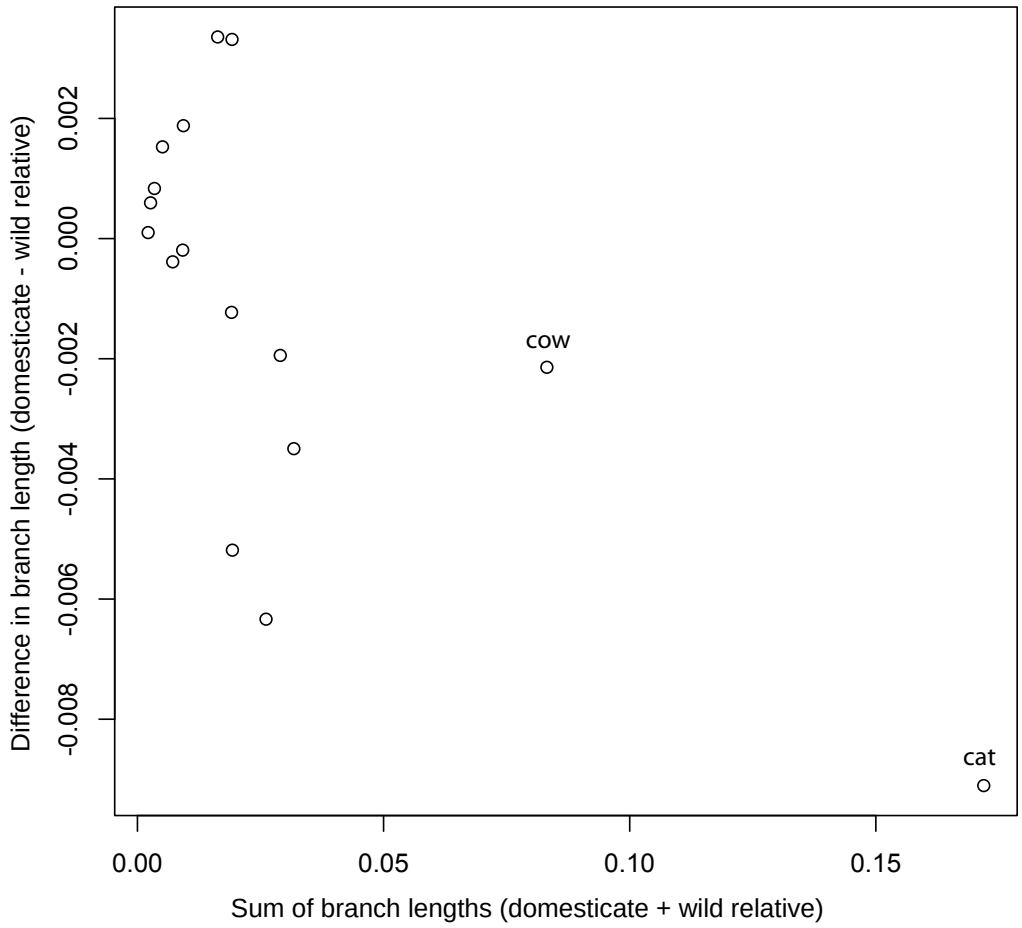
Table S2: Results of sister pairs analyses for subsets of the comparisons that exclude distant domesticate-wild relative comparisons or comparisons where the wild relatives have experienced a genetic bottleneck. The *p*-values from sign and Wilcoxon signed-ranks tests for the rate of non-synonymous substitutions (d_N), synonymous substitutions (d_S) and their ratio (d_N/d_S) along with total substitution rate from all 16 comparisons reported in the manuscript are compared with results when 1) the cat is removed (the most distant pair), 2) all distant comparisons are removed (cat, goat, cow, water buffalo, goose), 3) all comparisons where the wild relative (WR) has experienced a bottleneck are removed (horse, camel, pig, donkey, water buffalo), and 4) when both distant comparisons and comparisons with WR bottlenecks are removed. Significant results ($p \leq 0.05$) are indicated with an asterisk. We found $d_N = 0$ for one dog pair (see Table 1), so we repeated the Wilcoxon signed-ranks tests without this data point (values in parentheses).

Subset	Sign test *p*-values				Wilcoxon signed-ranks test *p*-values			
	d_N/d_S	**d_S**	**d_N**	**Branch length**	**d_N/d_S**	**d_S**	**d_N**	**Branch length**
All comparisons	1	0.45	1	0.80	0.75 (0.76)	0.78	1 (0.80)	0.32
Cat removed	1	0.30	1	1	0.85 (0.85)	0.45	1 (1)	0.52
Cat and cow removed	1	0.42	1	1	0.83 (0.84)	0.63	0.89 (0.89)	0.71
Distant comparisons removed	1	0.23	1	0.55	0.84 (0.85)	0.32	0.84 (0.85)	0.28
Comparisons with WR bottleneck removed	0.55	1	1	0.23	0.13 (0.13)	0.64	0.61 (0.62)	0.07
Comparisons with WR bottleneck removed (water buffalo also removed)	0.34	1	1	0.34	0.02* (0.02) *	0.49	0.24 (0.25)	0.13
Distant and WR bottleneck comparisons removed	0.69	0.69	1	1	0.11 (0.12)	1	0.59 (0.62)	1

Table S3: Results of whole tree analyses on several alternative subsets of the data. Results include d_N/d_S values for all mitochondrial coding genes for birds and mammals and branch lengths for complete mitochondrial genomes (excluding the D-loop) for birds and mammals. Results are also presented for subsets that exclude either 1) distantly related comparisons (cat, goat, cow, goose, water buffalo), 2) comparisons where there is evidence in the literature that wild relatives (WR) have undergone genetic bottlenecks (horse, camel, pig, donkey, water buffalo), or 3) both. All reported results for d_N/d_S and branch lengths were estimated in codeml and baseml in PAML v4.4b (see Methods). Results of one-rate and two-rate comparisons using the likelihood ratio test (LRT) are also reported (see Methods). We did not detect a significant rate difference between domesticate and wild branches in any of the whole tree analyses.

Birds and mammals - mt coding genes	**Domesticate d_N/d_S**	**Wild relative d_N/d_S**	**LRT *p*-value**
All comparisons	0.042	0.047	0.416
Cat and cow removed	0.056	0.055	0.926
Distant comparisons removed	0.056	0.051	0.689
Comparisons with WR bottlenecks removed	0.039	0.042	0.631
Comparisons with WR bottlenecks removed (water buffalo also removed)	0.039	0.037	0.720
Distant and wild relative bottleneck comparisons removed	0.062	0.032	0.093
Mammals - complete mt genomes	**Domesticate branch length**	**Wild relative branch length**	**LRT *p*-value**
All comparisons	1.049	1.045	0.946
Cat and cow removed	2.375	2.199	0.239
Distant comparisons removed	2.849	2.452	0.067
Comparisons with WR bottlenecks removed	0.867	0.949	0.141
Comparisons with WR bottlenecks removed (water buffalo also removed)	0.836	0.914	0.189
Distant and wild relative bottleneck comparisons removed	2.124	2.028	0.780

Figure S1: The relationship between divergence (domesticate branch length + wild relative branch length) and difference in branch length (domesticate – wild relative) for each sister pair. The most distant comparisons, the cat and cow, are highlighted.



Literature Cited

Agnarsson I, Kuntner M, May-Collado LJ. 2010. Dogs, cats, and kin: A molecular species-level phylogeny of Carnivora. Mol Phylogenet Evol. 54:726–745.

Amaral AJ, et al. 2011. Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. PLoS ONE 6:e14782.

Anthony DW, Brown DR. 1991. The origins of horseback riding. Antiquity 65:22–38.

Aranguren-Mendez J, Beja-Pereira A, Avellanet R, Dzama K, Jordana J. 2004. Mitochondrial DNA variation and genetic relationships in Spanish donkey breeds (*Equus asinus*). J Anim Breed Genet. 121:319–330.

Bailey J, et al. 1996. Ancient DNA suggests a recent expansion of European cattle from a diverse wild progenitor species. Proc R Soc Lond B Biol Sci. 263:1467–1473.

Beja-Pereira A, et al. 2004. African origins of the domestic donkey. Science 304:1781–1781.

Benirschke K, Malouf N, Low RJ, Heck H. 1965. Chromosome complement: differences between *Equus caballus* and *Equus przewalskii*, Poliakoff. Science 148:382–383.

Björnerfeldt S, Webster MT, Vilà C. 2006. Relaxation of selective constraint on dog mitochondrial DNA following domestication. Genome Res. 16:990–994.

Buckland RB, Guy G. 2002. *Goose Production*. Food and Agriculture Organisation of the United Nations, Rome, Italy.

Burton JA, Hedges S, Mustari AH. 2005. The taxonomic status, distribution and conservation of the lowland anoa *Bubalus depressicornis* and mountain anoa *Bubalus quarlesi*. Mammal review 35:25–50.

Brant AW. 1998. A brief history of the turkey. World's Poultry Science Journal 54:365–373.

Chessa B, et al. 2009. Revealing the history of sheep domestication using retrovirus integrations. Science 324:532–536.

Clutton-Brock J. 1999. *A natural history of domesticated mammals*. 2nd ed. Cambridge University Press, Cambridge, U. K.

Crawford RD. 1992. Introduction to Europe and diffusion of domesticated turkeys from the America. Arch Zootec. 41:307–314.

Darwin C. 1868. *The variation of animals and plants under domestication*. 1st ed. John Murray, London, U. K.

Donne-Gousse C, Laudet V, Hanni C. 2002. A molecular phylogeny of Anseriformes based on mitochondrial DNA analysis. Mol Phylogenet Evol. 23:339–356.

Driscoll CA, Macdonald DW, O'Brien SJ. 2009. From wild animals to domestic pets, an evolutionary view of domestication. Proc Natl Acad Sci U S A. 106:9971–9978.

Driscoll CA, et al. 2007. The Near Eastern origin of cat domestication. Science 317:519–523.

Druzhkova AS et al. 2013. Ancient DNA analysis affirms the Canid from Altai as a primitive dog. PLoS ONE 8:e57754.

Fumihito A, et al. 1996. Monophyletic origin and unique dispersal patterns of domestic fowls. Proc Natl Acad Sci U S A. 93:6792–6795.

Fumihito A, et al. 1994. One subspecies of the red junglefowl (*Gallus gallus gallus*) suffices as the matriarchic ancestor of all domestic breeds. Proc Natl Acad Sci U S A. 91:12505–12509.

Giuffra E, et al. 2000. The origin of the domestic pig: Independent domestication and subsequent introgression. Genetics 154:1785–1791.

Goto H, et al. 2011. A massively parallel sequencing approach uncovers ancient origins and high genetic variability of endangered Przewalski's horses. Genome Biol Evol. 3:1096–1106.

Guo S, et al. 2006. Origin of mitochondrial DNA diversity of domestic yaks. BMC Evol Biol. 6:73.

Hare J. 1997. The wild Bactrian camel *Camelus bactrianus ferus* in China: The need for urgent action. Oryx 31:45–48.

Harris D. 1962. The distribution and ancestry of the domestic goat. Biol J Linn Soc. 173:79–91.

Hassanin A, Ropiquet A. 2004. Molecular phylogeny of the tribe Bovini (Bovidae, Bovinae) and the taxonomic status of the Kouprey, *Bos sauveli* Urbain 1937. Mol Phylogenet Evol. 33:896–907.

Hassanin A, Ropiquet A, Couloux A, Cruaud C. 2009. Evolution of the mitochondrial genome in mammals living at high altitude: New insights from a study of the tribe Caprini (Bovidae, Antilopinae). J Mol Evol. 68:293–310.

Hassanin A, Bonillo C, Nguyen BX, Cruaud C. 2010. Comparisons between mitochondrial genomes of domestic goat (*Capra hircus*) reveal the presence of numts and multiple sequencing errors. Mitochondrial DNA 21:68–76.

Hemmer H. 1990. *Domestication: The decline of environmental appreciation*, trans. N. Beckhaus. Cambridge University Press, Cambridge, U. K.

Hiendleder S, Kaupe B, Wassmuth R, Janke A. 2002. Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. Proc R Soc Lond B Biol Sci. 269:893–904.

Hiendleder S, Lewalski H, Janke A. 2008. Complete mitochondrial genomes of *Bos taurus* and *Bos indicus* provide new insights into intra-species variation, taxonomy and domestication. Cytogenet. Genome Res. 120:150–156.

Hiendleder S, Mainz K, Plante Y, Lewalski H. 1998. Analysis of mitochondrial DNA indicates that domestic sheep are derived from two different ancestral maternal sources:

No evidence for contributions from urial and argali sheep. J Hered. 89:113-120.

Ji R, et al. 2009. Monophyletic origin of domestic Bactrian camel (*Camelus bactrianus*) and its evolutionary relationship with the extant wild camel (*Camelus bactrianus ferus*). Anim Genet. 40:377–382.

Kadwell M, et al. 2001. Genetic analysis reveals the wild ancestors of the llama and the alpaca. Proc R Soc Lond B Biol Sci. 268:2575–2584.

Kavar T, Dovč P. 2008. Domestication of the horse: Genetic relationships between domestic and wild horses. Livestock Science 116:1–14.

Kierstein G, et al. 2004. Analysis of mitochondrial D-loop region casts new light on domestic water buffalo (*Bubalus bubalis*) phylogeny. Mol Phylogenet Evol. 30:308–324.

Kijas JM, Andersson L. 2001. A phylogenetic study of the origin of the domestic pig estimated from the near-complete mtDNA genome. J Mol Evol. 52:302–308.

Kikkawa Y, Yonekawa H, Suzuki H. 1997. Analysis of genetic diversity of domestic water buffaloes and anoas based on variations in the mitochondrial gene for cytochrome b. Anim Genet. 28:195-201.

Kimball RT, Braun EL. 2008. A multigene phylogeny of Galliformes supports a single origin of erectile ability in non-feathered facial traits. J Avian Biol. 39:438–445.

Kimura B, et al. 2010. Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. Proc R Soc Lond B Biol Sci. 278:50–57.

Larson G. 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science 307:1618–1621.

Levine MA. 1999. Botai and the origins of horse domestication. J Anthropol Archaeol. 18:29–78.

Li SP, et al. 2007. Molecular phylogeny and taxonomic status of domestic yak inferred from cytochrome b gene partial sequences. J Anim Vet Adv. 6:1495–1499.

Loftus RT, MacHugh DE, Bradley DG, Sharp PM, Cunningham P. 1994. Evidence for two independent domestications of cattle. Proc Natl Acad Sci. USA 91:2757–2761.

Luikart G, et al. 2001. Multiple maternal origins and weak phylogeographic structure in domestic goats. Proc Natl Acad Sci U S A. 98:5927–5932.

MacEachern S, et al. 2009. Molecular evolution of the Bovini tribe (Bovidae, Bovinae): Is there evidence of rapid evolution or reduced selective constraint in domestic cattle? BMC Genomics 10:179.

Marín JC, et al. 2007a. Mitochondrial phylogeography and demographic history of the vicuña: Implications for conservation. Heredity 99:70–80.

Marín JC, et al. 2007b. Systematics, taxonomy and domestication of alpaca and llama:

New chromosomal and molecular evidence. Rev Chil Hist Nat. 80:121–140.

Meadows JRS, Cemal I, Karaca O, Gootwine E, Kijas JW. 2006. Five ovine mitochondrial lineages identified from sheep breeds of the Near East. Genetics 175:1371–1379.

Meadows JRS, Hiendleder S, Kijas JW. 2010. Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. Heredity 106:700–706.

Moehlman PD. 2002. Status and action plan for the African wild ass. In: Moehlman P. D., editor. *Equids: zebras, asses and horses, status survey and conservation action plan*. Gland, Switzerland: IUCN. p. 2–9.

Nishibori M, Shimogiri T, Hayashi T, Yasue H. 2005. Molecular evidence for hybridization of species in the genus *Gallus* except for *Gallus varius*. Anim Genet. 36:367–375.

Ollivier L. 2009. European pig genetic diversity: A minireview. Animal 3:915–924.

Pedrosa S, et al. 2005. Evidence of three maternal lineages in near eastern sheep supporting multiple domestication events. Proc R Soc Lond B Biol Sci. 272:2211–2217.

Pertoldi C, et al. 2010. Phylogenetic relationships among the European and American bison and seven cattle breeds reconstructed using the BovineSNP50 Illumina Genotyping BeadChip. Acta Theriol. 55:97–108.

Pidancier N, Jordan S, Luikart G, Taberlet P. 2006. Evolutionary history of the genus *Capra* (Mammalia, Artiodactyla): Discordance between mitochondrial DNA and Y-chromosome phylogenies. Mol Phylogenet Evol. 40:739–749.

Ramírez O, et al. 2009. Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds. Mol Biol Evol. 26:2061–2072.

Rhode D, Madsen DB, Brantingham PJ, Dargye T. 2007. Yaks, yak dung, and prehistoric human habitation of the Tibetan Plateau. Dev Quat Sci. 9:205–224.

Robinson TJ, Ropiquet A. 2011. Examination of hemiplasy, homoplasy and phylogenetic discordance in chromosomal evolution of the Bovidae. Syst Biol. 60:439–450.

Rossel S, et al. 2008. Domestication of the donkey: Timing, processes, and indicators. Proc Natl Acad Sci U S A. 105:3715–3720.

Ruokonen M, Kvist L, Lumme J. 2000. Close relatedness between mitochondrial DNA from seven *Anser* goose species. J Evol Biol. 13:532–540.

Ryder OA. 1993. Przewalski's horse: Prospects for reintroduction into the wild. Conserv Biol. 7:13–15.

Sasaki M, et al. 1999. Morphological character of the shoulder and leg skeleton in Przewalski's horse (*Equus przewalskii*). Ann Anat. 181:403–407.

Savolainen P, Zhang YP, Luo J, Lundeberg J, Leitner T. 2002. Genetic evidence for an East Asian origin of domestic dogs. Science 298:1610–1613.

Sawai H, et al. 2010. The origin and genetic variation of domestic chickens with special reference to junglefowls *Gallus g. gallus* and *G. varius*. PLoS ONE 5:e10639.

Scandura M, et al. 2008. Ancient vs. recent processes as factors shaping the genetic variation of the European wild boar: Are the effects of the last glaciation still detectable? Mol Ecol. 17:1745–1762.

Shen YY, et al. 2010. A mitogenomic perspective on the ancient, rapid radiation in the Galliformes with an emphasis on the Phasianidae. BMC Evol Biol. 10:132.

Shi XW, Wang JW, Zeng FT, Qiu XP. 2006. Mitochondrial DNA cleavage patterns distinguish independent origin of Chinese domestic geese and Western domestic geese. Biochem Genet. 44:237–245.

Shinan R. 1996. Several major achievements in early Neolithic China, ca. 5000 BC, transl. W. K. Cheung, ed. B. Gordon. Kaogu−Archaeology 37–49.

Silbermayr K, et al. 2010. High mitochondrial differentiation levels between wild and domestic Bactrian camels: A basis for rapid detection of maternal hybridization. Anim Genet. 41:315–318.

Speller CF, et al. 2010. Ancient mitochondrial DNA analysis reveals complexity of indigenous North American turkey domestication. Proc Natl Acad Sci U S A. 107:2807–2812.

Takada T, Kikkawa Y, Yonekawa H, Kawakami S, Amano T. 1997. Bezoar (*Capra aegagrus*) is a matriarchal candidate for ancestor of domestic goat (*Capra hircus*): Evidence from the mitochondrial DNA diversity. Biochem Genet. 35:315–326.

Verkaar ELC. 2004. Maternal and paternal lineages in cross-breeding Bovine species. Has wisent a hybrid origin? Mol Biol Evol. 21:1165–1170.

Vigne J, Guilaine J, Debue K, Haye L, Gérard P. 2004. Early taming of the cat in Cyprus. Science 304:259.

Vilà C, et al. 2001. Widespread origins of domestic horse lineages. Science 291:474–477.

Vilà C, et al. 1997. Multiple and ancient origins of the domestic dog. Science 276:1687–1689.

Wade CM, et al. 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science 326:865–867.

Wallner B, Brem G, Müller M, Achmann R. 2003. Fixed nucleotide differences on the Y chromosome indicate clear divergence between *Equus przewalskii* and *Equus caballus*. Anim Genet. 34:453–456.

Wang CM, et al. 2010a. The origin of the white Roman goose. Biochem Genet. 48:938–943.

Wang Z, et al. 2011. Domestication relaxed selective constraints on the yak mitochondrial genome. Mol Biol Evol. 28:1553–1556.

Wang Z, et al. 2010b. Phylogeographical analyses of domestic and wild yaks based on mitochondrial DNA: New data and reappraisal. J Biogeogr. 37:2332–2344.

Wheeler JC. 1995. Evolution and present situation of the South American Camelidae. Biol J Linn Soc Lond. 54:271–295.

Wheeler JC, Russel A, Redden H. 1995. Llamas and alpacas: Pre-conquest breeds and post-conquest hybrids. J Archaeol Sci. 22:833–840.

Wiener G, Jianlin H, Ruijun L. 2003. *The yak*. 2nd ed. The Regional Office for Asia and the Pacific of the Food and Agriculture Organization of the United Nations, Bangkok, Thailand.

Woolfit M, Bromham L. 2005. Population size and molecular evolution on islands. Proc R Soc Lond B Biol Sci. 272:2277-2282.

Wu GS, et al. 2007. Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multiple domestication events in East Asia. Genome Biol. 8:R245.

Xu X, Gullberg A, Arnason U. 1996. The complete mitochondrial DNA (mtDNA) of the donkey and mtDNA comparisons among four closely related mammalian species-pairs. J Mol Evol. 43:438–446.

Yindee M, et al. 2010. Y-chromosomal variation confirms independent domestications of swamp and river buffalo. Anim Genet. 41:433-435.

Zeder MA. 2000. The initial domestication of goats (*Capra hircus*) in the Zagros Mountains 10,000 years ago. Science 287:2254–2257.

Zhao S, et al. 2012. Molecular phylogeny of major lineages of the avian family Phasianidae inferred from complete mitochondrial genome sequences. J. Nat. Hist. 46:757–767.

Zhu W, et al. 2010. Two maternal origins of the Chinese domestic grey goose. J. Anim Vet Adv. 9:2674–2678.

Chapter 6

Conclusions

In my thesis I used a macroevolutionary approach to investigate several questions related to the evolution of salt tolerance in angiosperms. The use of phylogenetic comparative methods allowed me to utilize the tremendous amount of research on halophytes and the wealth of published genetic and phylogenetic data cataloging the diversity of angiosperms to identify several remarkable facets of the evolution of salt tolerance.

Despite the rarity and complexity of salt tolerance, we find that it has evolved hundreds of times in the angiosperms. This finding seems counterintuitive in that salt tolerance often involves many physiological or anatomical modifications. Furthermore, if salt tolerance has evolved repeatedly in some lineages such as the Poaceae, which contains many major crop species, why has it been difficult to breed salt tolerant crops?

One explanation is that salt tolerance may be costly to develop and maintain. This cost can take many forms, for example both increasing ion selectivity and exclusion and producing compatible solutes and osmoprotectants have energetic costs that can reduce growth rates (Yeo 1983; Cheeseman 1988; Flowers and Yeo 1995; Eallonardo et al. 2012). Oxidative stress could also impose a perhaps unavoidable cost on salt tolerant species: The repeated or prolonged exposure to the toxic effects of reactive oxygen species (ROS) induced by salinity (Mittler 2002; Bose et al. 2014) could lead to the accumulation of mutations (Jiang et al. 2014).

The finding that halophytes have increased molecular substitution rates in mitochondrial genes (Chapter 4), supports the existence of an oxidative cost of exposure to saline habitats, as oxidative stress may increase mutation rates, especially in the mitochondria (Mittler 2002; Bose et al. 2014). However, further investigation into the role of ROS in abiotic stress is required, as evidence suggests that while ROS may be damaging, they are also crucial in stress signaling, including the activation of antioxidant and osmoprotective pathways (Apel and Hirt 2004; Fujita et al. 2006; Rhoads 2006; Miller et al. 2008; Mittler et al. 2011). Furthermore, it is clear that many halophytes have developed mechanisms to resist oxidative damage (Lechno et al. 1997; Ozgur et al. 2013; Bose et al. 2014), and so the signal of increased substitution rates in mitochondrial genes may be due to another aspect of the biology or life history of halophytes rather than exposure to ROS.

The processes of how non-salt tolerant lineages develop salt tolerance and transition to saline habitats is an open question in our understanding of the evolution of salt tolerance and the biodiversity patterns of halophytes. It is unclear how the permanence of saline habitats has shaped the evolution of halophytes since some saline areas are long lived, while others are seasonal or ephemeral on evolutionary time scales (Ungar 1998; Bui 2013). Could the diversity in the spatial and temporal patterns of salinity drive or explain the diversity of the evolutionary patterns we observe in angiosperms? Perhaps long-standing salt lakes and salt marshes generally support halophytes from families with conserved salt tolerance origins, for example, the Amaranthaceae (Kadereit et al. 2012), while more labile evolutionary patterns are shaped by the transient nature of land salinity in other habitats. One way to investigate this question is to incorporate spatial data into analyses using a comparative community phylogenetics approach. For example, comparing the spatial and genetic patterns of halophytes with non-salt tolerant species within verses around saline areas (e.g., many comparisons of species in a salt marsh and species in the land surrounding the salt marsh) would be valuable for integrating spatial data with broader phylogenetic information on the biodiversity and evolutionary patterns of salt tolerance. A community phylogenetic approach may answer questions relating to the colonization of and relationship between saline habitats like: 1) What is the relationship between species in highly salt tolerant areas across a diverse geographic scale? 2) Are species in highly saline areas more closely related to species in other highly saline areas or to the species surrounding the local saline habitat?

We also found evidence that another rare and complex ecophysiological trait, heavy metal hyperaccumulation, appears to have evolved repeatedly among angiosperm families, often among closely related species. Furthermore, we find evidence that halophytes and heavy metal hyperaccumulators are significantly associated taxonomically among angiosperm families and significantly related among species in some families. These observations provide evidence that these abilities may rely on similar physiological mechanisms (Stewart and Lee 1974; Schat et al. 1997; Manousaki and Kalogerakis 2011), and that some lineages are more likely to produce both halophytes and heavy metal hyperaccumulators than others. For example, there is some evidence that two traits associated with salt tolerance, leaf mass area and leaf nitrogen (Eallonardo et al., 2012), are evolutionarily labile among *Helianthus* species and populations of the hyperaccumulator *Helianthus annuus* (Donovan et al., 2014). So if a lineage has the underlying physiological traits that contribute to both conditions, a

species from that lineage may be more likely to successfully move into a novel harsh environment, developing either salt tolerance or hyperaccumulation (Cheeseman, 2014).

This association, and the association between abiotic tolerances more broadly, could also be driven by the overlap in stress signaling pathways associated with abiotic stimuli, allowing some lineages to be tolerant to multiple abiotic stresses (Fujita et al. 2006; Mittler et al. 2011; Golldack et al. 2014). We believe this study system will provide an important backdrop for finer scale studies on the mechanisms that support and drive the relationship between these abilities, as well as the connection between abiotic stress tolerances more broadly. We also produce a list of species known as both halophytes and heavy metal hyperaccumulators, which we believe will be useful in future studies on multiple tolerance and in the identification and development of efficient (e.g., fast growing, highly tolerant) species for practical use (Qadir et al. 2007; Manousaki and Kalogerakis 2011; Lutts and Lefevre 2015).

During my thesis, I also had the opportunity to contribute to a study exploring the relationship between salt tolerance and another abiotic stress tolerance, alkalinity tolerance. In this study we explored alternative techniques to list-based approaches for identifying tolerant species and exploring the connection between abiotic stress tolerances. This study employed geochemical modeling of soil samples and spatial data on species distributions to explore the relationship between salinity and alkalinity tolerance in the Australian Poaceae (Saslis-Lagoudakis et al. 2015, Appendix I). We believe the techniques used in this study offer a promising direction in capturing and understanding the continuity of stress tolerance as it relates to the geographical distribution of species. This method may also prove useful in analyzing the patterns of halophyte diversity and investigating the colonization of saline habitats, as described above.

In Chapter 5, we investigated the generality of several case studies (Björnerfeldt et al. 2006; Cruz et al. 2008; Wang et al. 2011; Hughes 2013) reporting increased rates of molecular evolution in domesticated animals compared to their wild relatives. We found no evidence that domesticated mammals and birds have consistently higher rates of molecular evolution compared to wild relatives in the mitochondrial genome. In support of this finding, one recent study reports that several mammal domesticates do not have higher recombination rates compared to wild relatives (Munoz-Fuentes et al. 2015). If

this is true across a broad range of domesticates, this could explain why domestication may not consistently increase rates of molecular evolution. However, domestication clearly leads to incredible changes over evolutionarily short time scales, and so the identification of the consistent mechanisms that lead to and maintain these changes is still an intriguing area of research. In future it would be interesting to repeat the analysis we conducted in plants to see if the effect of domestication is consistent between plants and animals. If domestication does not consistently influence rates of molecular evolution in plants, this could provide evidence that difficulties in breeding salt tolerance may be less attributable to genomic changes associated with domestication than to other aspects of the evolution of salt tolerance.

References

Apel, K., and H. Hirt. 2004. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annual Review of Plant Biology 55:373–399.

Björnerfeldt, S., M. T. Webster, and C. Vilà. 2006. Relaxation of selective constraint on dog mitochondrial DNA following domestication. Genome Research 16:990–994.

Bose, J., A. Rodrigo-Moreno, and S. Shabala. 2014. ROS homeostasis in halophytes in the context of salinity stress tolerance. Journal of Experimental Botany 65:1241–1257.

Bui, E. N. 2013. Soil salinity: A neglected factor in plant ecology and biogeography. Journal of Arid Environments 92:14–25.

Cheeseman, J. M. 1988. Mechanisms of salinity tolerance in plants. Plant Physiology 87:547–550.

Cruz, F., C. Vilà, and M. T. Webster. 2008. The legacy of domestication: accumulation of deleterious mutations in the dog genome. Molecular Biology and Evolution 25:2331–2336.

Donovan, L. A., Mason, C. M., Bowsher, A. W., Goolsby, E. W. & Ishibashi, C. D. A. 2014. Ecological and evolutionary lability of plant traits affecting carbon and nutrient cycling. Journal of Ecology 102:302–314.

Eallonardo, A. S., Jr, D. J. Leopold, J. D. Fridley, and J. C. Stella. 2012. Salinity tolerance and the decoupling of resource axis plant traits. Journal of Vegetation Science 24:365–374.

Flowers, T. J., and A. R. Yeo. 1995. Breeding for salinity resistance in crop plants: where next? Australian Journal of Plant Physiology 22:875–884.

Fujita, M., Y. Fujita, Y. Noutoshi, F. Takahashi, Y. Narusaka, K. Yamaguchi-Shinozaki, and K. Shinozaki. 2006. Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. Current Opinion in Plant Biology 9:436–442.

Golldack, D., Li, C., Mohan, H., and N. Probst. 2014. Tolerance to drought and salt stress in plants: unraveling the signaling networks. Frontiers in Plant Science 5:151.

Hughes, A. L. 2013. Accumulation of slightly deleterious mutations in the mitochondrial genome: a hallmark of animal domestication. Gene 515:28–33.

Jiang, C., A. Mithani, E. J. Belfield, R. Mott, L. D. Hurst, and N. P. Harberd. 2014. Environmentally responsive genome-wide accumulation of de novo Arabidopsis thalianamutations and epimutations. Genome Research 24:1821–1829.

Kadereit, G., D. Ackerly, and M. D. Pirie. 2012. A broader model for C_4 photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae s.s.). Proceedings of the Royal Society of London B: Biological Sciences 279:3304–3311.

Lechno, S., E. Zamski, and E. Tel-Or. 1997. Salt stress-induced responses. Journal of Plant Physiology 150:206–211.

Lutts, S., and I. Lefevre. 2015. How can we take advantage of halophyte properties to cope with heavy metal toxicity in salt-affected areas? Annals of Botany 0:1–20.

Manousaki, E., and N. Kalogerakis. 2011. Halophytes—an emerging trend in phytoremediation. International Journal of Phytoremediation 13:959–969.

Miller, G., V. Shulaev, and R. Mittler. 2008. Reactive oxygen signaling and abiotic stress. Physiologia Plantarum 133:481–489.

Mittler, R. 2002. Oxidative stress, antioxidants and stress tolerance. Trends in Plant Science 7:405–410.

Mittler, R., S. Vanderauwera, N. Suzuki, G. Miller, V. B. Tognetti, K. Vandepoele, M. Gollery, V. Shulaev, and F. Van Breusegem. 2011. ROS signaling: the new wave? Trends in Plant Science 16:300–309.

Munoz-Fuentes, V., M. Marcet-Ortega, G. Alkorta-Aranburu, C. Linde Forsberg, J. M. Morrell, E. Manzano-Piedras, A. Soderberg, K. Daniel, A. Villalba, A. Toth, A. Di Rienzo, I. Roig, and C. Vilà. 2015. Strong artificial selection in domestic mammals did not result in an increased recombination rate. Molecular Biology and Evolution 32:510–523.

Ozgur, R., B. Uzilday, A. H. Sekmen, and I. Turkan. 2013. Reactive oxygen species regulation and antioxidant defence in halophytes. Functional Plant Biology 40:832–847.

Qadir, M., J. D. Oster, S. Schubert, A. D. Noble, and K. L. Sahrawat. 2007. Phytoremediation of sodic and saline-sodic soils. Advances in Agronomy 96:197–247.

Rhoads, D. M. 2006. Mitochondrial reactive oxygen species. Contribution to oxidative stress and interorganellar signaling. Plant Physiology 141:357–366.

Saslis-Lagoudakis, C. H., X. Hua, E. Bui, C. Moray, and L. Bromham. 2015. Predicting species' tolerance to salinity and alkalinity using distribution data and geochemical modelling: a case study using Australian grasses. Annals of Botany 115:343–351.

Schat, H., S. S. Sharma, and R. Vooijs. 1997. Heavy metal-induced accumulation of free proline in a metal-tolerant and a nontolerant ecotype of *Silene vulgaris*. Physiologia Plantarum 101:477–482.

Stewart, G. R., and J. A. Lee. 1974. The role of proline accumulation in halophytes. Planta 120:279–289.

Ungar, I. A. 1998. Are biotic factors significant in influencing the distribution of halophytes in saline habitats? The Botanical Review 64:176–199.

Wang, Z., T. Yonezawa, B. Liu, T. Ma, X. Shen, J. Su, S. Guo, M. Hasegawa, and J. Liu. 2011. Domestication relaxed selective constraints on the yak mitochondrial genome. Molecular Biology and Evolution 28:1553–1556.

Yeo, A. R. 1983. Salinity resistance: physiologies and prices. Physiologia Plantarum 58:214–222.

Page intentionally left blank

Appendix I

Predicting species' tolerance to salinity and alkalinity using distribution data and geochemical modelling: a case study using Australian grasses

C. Haris Saslis-Lagoudakis[1], Xia Hua[1], Elisabeth Bui[2], **Camile Moray**[1], and Lindell Bromham[1]

[1] Centre for Macroevolution and Macroecology, Research School of Biology, Australian National University, Canberra, Australian Capital Territory 0200, Australia
[2] CSIRO Land and Water, GPO Box 1666, Canberra, Australian Capital Territory 2601, Australia

Published: Annals of Botany. 2015. 115:343-351

Abstract

Background and Aim: Salt tolerance has evolved many times independently in different plant groups. One possible explanation for this pattern is that it builds upon a general suite of stress-tolerance traits. If this is the case, then we might expect a correlation between salt tolerance and other tolerances to different environmental stresses. This association has been hypothesised for salt and alkalinity-tolerance. However, a major limitation in investigating large-scale patterns of these tolerances is that lists of known tolerant species are incomplete. Here, we explore whether we can predict species' salt and alkalinity-tolerance using geochemical modelling for Australian grasses. Then, we assess the correlation between taxa found in conditions of high predicted salinity and alkalinity.

Methods: We use extensive occurrence data for Australian grasses and geochemical modelling to predict values of pH and electrical conductivity (EC) to which species are exposed in their natural distributions. Using parametric and phylogeny-corrected tests, we i) evaluate our geochemical predictions using a list of known halophytes as a control, and ii) ask whether taxa that occur in conditions of high predicted salinity are also found in conditions of high predicted alkalinity.

Key Results: We show that genera containing known halophytes have higher predicted salinity conditions than those not containing known halophytes. Additionally, we find that taxa occurring in high predicted salinity tend to also occur in high predicted alkalinity.

Conclusions: Geochemical modelling using species' occurrence data is a potentially useful approach to predict species' relative natural tolerance to challenging environmental conditions. Our findings also demonstrate a correlation between salinity and alkalinity-tolerance. Further investigations can consider the phylogenetic distribution of specific traits involved in these ecophysiological strategies, ideally by incorporating more complete finer scale geochemical information, as well as laboratory experiments.

Keywords: alkalinity-tolerance, geochemical modelling, macroevolution, phylogeny, Poaceae, salt tolerance, stress resistance syndrome

Introduction

Many plant species have developed several ecophysiological strategies to tolerate extreme conditions in challenging environments. For example, species that complete their life cycle in saline environments - known as halophytes - have evolved various mechanisms that have enabled them to survive and reproduce in these environments (Flowers & Colmer, 2008; Munns & Tester, 2008). These mechanisms are related to water uptake and defence against ion toxicity within the plant, such as the accumulation and compartmentalisation of saline ions, the ability to limit the entry of these ions into the transpiration stream, the synthesis of compatible solutes for osmoprotection, the ability to accumulate essential nutrients, and the ability to continue to regulate transpiration in the presence of high concentrations of Na^+ and Cl^- (Deinlein et al., 2014; Flowers & Colmer, 2008; Munns & Tester, 2008; Rozema & Flowers, 2008; Shabala, 2013). Research has unveiled the complex, physiological, molecular, and genetic background of these adaptations [e.g. (Ashraf & Foolad, 2013; Munns, 2005; Munns & Tester, 2008; Shavrukov, 2012)]. There are more than 1,500 species of halophytes (Aronson, 1989) and salt tolerance is widely distributed across the plant phylogeny, with multiple independent origins (Flowers et al., 1977; Saslis-Lagoudakis et al., 2014). However some plant groups, such as Caryophyllales and Alismatales, contain more halophytes than others (Flowers et al., 2010; Saslis-Lagoudakis et al., 2014). At a lower hierarchical level, salt tolerance has also evolved multiple times independently. For example, it has evolved over 70 times in the grass family alone, and is phylogenetically non-random, i.e. some clades are more likely than others to contain salt tolerant species (Bennett et al., 2013).

It has been suggested that tolerance mechanisms and physiological responses to salinity are shared with other types of environmental stresses, such as aridity, flooding, and frost (Munns & Tester, 2008; Rozema & Schat, 2013; Tuteja, 2007). For example, a recent study found that salt tolerance in grasses evolves more frequently in C_4 than C_3 lineages, demonstrating a close association in the evolution of C_4 photosynthesis and salt tolerance in these lineages (Bromham & Bennett, 2014). This type of correlations may provide one possible explanation for the repeated evolution of salt tolerance. The stress resistance syndrome hypothesis (Chapin et al., 1993) states that there may be a suite of stress-related traits that allow plants to survive in a variety of stressful environments.

Therefore, the presence of "enablers" in some lineages can facilitate the evolution of multiple stress resistance within those lineages (Edwards & Donoghue, 2013). This suggests that traits related to tolerance to one type of stress can facilitate the evolution of another type of stress-resistance. For example, salt tolerance, succulence and C_4 photosynthesis are associated in chenopods (Kadereit et al., 2012) and occupation of bare environments served as an "enabler" to adaptation to harsh elemental soils in the Brassicaceae (Cacho & Strauss, 2014). Therefore, by studying these ecophysiological traits in a phylogenetic context, we can investigate macroevolutionary patterns of ecophysiological evolution (Ackerly et al., 2000), and explore the correlation between different ecophysiological strategies (Niinemets & Valladares, 2006).

A correlation of this kind has been suggested between salt and alkaline tolerance (Bromham et al., 2013; Bui, 2013; Bui et al., 2014). Alkalinity (high soil pH) often co-occurs with salinity (high soil NaCl concentrations) in the landscape: many saline soils are also alkaline due to the presence of sodium carbonates (Rengasamy, 2010). Therefore, it is possible that lineages occupying these environments have had to evolve strategies to cope with both alkalinity and salt-stress (Bui, 2013). Like salinity, alkalinity exacerbates water loss, interfering with stomatal closure due to the accumulation of sodium ions (Bernstein, 1975). Soils of high pH often have poor structure, affecting their hydraulic conductivity and the plants' water uptake, and causing hypoxia in the root zone (Bernstein, 1975). Both these factors affect water use efficiency, which is also one of the major stresses for plants in saline environments. Plants equipped to deal with salinity and alkalinity employ osmotic adjustments (Farrell et al., 1996; Yang et al., 2007; Yang et al., 2008) that are not found in plants without tolerance to either stresses (Chen et al., 2011; Liu et al., 2010), and which make tolerant plants naturally resistant to water stress (García & Mendoza, 2014). Further, both salinity and alkalinity affect photosynthesis and metabolism through a range of physiological and molecular processes (Nishiuchi et al., 2010; Yang et al., 2008). It is possible, therefore, that because of the shared challenges, salt and alkaline tolerance have evolved in closely related lineages which possess traits enabling the evolution of mechanisms of tolerance to either stress.

One of the main constraints in exploring large-scale patterns in salt and alkaline tolerance is the lack of exhaustive published lists of halophytes and particularly alkaline-tolerant species. Because field and laboratory observations of plant species'

tolerance to salinity and alkalinity tend to focus on particular species, lists of known halophytes are likely to be incomplete, and there are no comprehensive lists of alkaline tolerant species. An alternative approach to generating such lists is to predict plant species that are tolerant to these stresses based on their geographical distributions. In the last two decades, inferring species' environmental niche preferences from their natural distributions and environmental GIS data layers has become commonplace in studies of ecology and evolution (Guisan & Thuiller, 2005; Kozak et al., 2008; Warren et al., 2008). By combining distribution data with geochemical observations, we can infer salinity and alkalinity conditions to which species are exposed in their natural distributions. Although microbial studies have combined geochemical data with phylogenetic metrics (Costa et al., 2009; Macur et al., 2004; Reysenbach & Shock, 2002), geochemical modelling has been largely overlooked in studies of macroecology and macroevolution. However, a recent phylogenetic study of Australian *Acacia* species used geochemical modelling to investigate evolutionary patters of salinity and alkalinity tolerance (Bui et al., 2014).

The aims of this study were twofold: i) to evaluate the performance of geochemical modelling using species occurrence data, to identify species' tolerance to salinity and alkalinity, and ii) to investigate the correlation between salt and alkaline tolerance. We use Australian grasses (Poaceae) as a test case, because they are a group with a continent-wide distribution, occupying a wide range of environmental conditions, including arid, saline and sodic environments. Our dataset included distribution data for 1,387 species of mainland Australian grasses, of which 141 are known halophytes.

Materials and Methods

We investigated whether we could predict species' salt and alkaline tolerance based on species distribution modelling. To do that, we used geochemical modelling to generate species' descriptors for electrical conductivity (EC) and pH at their natural distributions. We evaluated the prediction of salt-tolerant species based on prior knowledge of salt tolerance in Australian grasses. Subsequently, we tested for the correlation between salt and alkalinity tolerance, and we explored if spatial patterns can explain this association.

In the literature, salinity and alkalinity tolerance are often characterised based on EC and pH soil values, respectively. For example, soils with EC over 4,000 μS m^{-1} are characterised as saline (United States Salinity Laboratory Staff, 1969) and plants tolerating 8,000 μS m^{-1} or over are considered halophytes (Aronson, 1989). Similarly, soil pH of 7 or higher is alkaline and most plants prefer pH 5.5–6.5 (Islam et al., 1980). In this study, we do no apply a threshold of EC or pH to characterise soils as saline or alkaline. Instead, we perform a comparative analysis of EC and pH conditions to which Australian grasses are exposed.

Predicting salt and alkalinity tolerance from species distribution modelling

Predicting species salt and alkalinity tolerance from occurrence data

Because there are no exhaustive databases that describe tolerances of all Australian grasses to salinity and alkalinity, in order to estimate these tolerances we employed an approach based on species' distributions. Our approach assumes that conditions of salinity and alkalinity at localities at which species are found naturally reflect their levels of tolerance to these conditions. Although factors other than tolerance affect species' distributions, such as interspecific competition, we can expect intrinsic tolerance to be correlated with realised tolerances. Therefore, it is possible to describe species' tolerances if we know: i) species' distributions and, ii) levels of salinity and alkalinity in these distributions. To generate species' distributions, we extracted occurrence data from the Atlas of Living Australia (ALA), a continent-wide dataset that contains approximately 45 million occurrence records for Australian biodiversity. There are 1,387 grass species found in mainland Australia (excluding Tasmania and other islands). Australian grass species are recorded from 354,913 points with unique geographic coordinates in the Atlas of Living Australia. We extracted all unique occurrence points for each species and we consider the distribution of each species to be the compilation of all the points at which it is reported.

In order to infer soil pH and electrical conductivity (EC) at the localities where grass species were reported, we accessed data from the National Geochemical Survey of Australia. This dataset reports the pH and EC on 1:5 soil:water extracts from bulk samples at 1,315 georeferenced point measurements across the continent, with an average sample density of 1 site/5,500 km^2 (de Caritat & Cooper, 2011). We retrieved

indications of EC and pH from the dataset and performed the analyses described below for subsoil (60-80 cm below the surface). Subsoil indications of EC and pH are more likely to reflect tolerance to salinity and alkalinity than shallower samples, as root tips - generally found deeper in the soil - are more highly sensitive to geochemistry than the rest of the root (Shabala, 2013).

From this dataset of subsoil EC and pH indications, we estimated EC and pH at each locality with a reported grass occurrence using Geostatistics in *geoR* (Diggle & Ribeiro, 2007). Geostatistics are techniques for mapping of surfaces from limited sample data and the estimation of values at unsampled locations in two steps (Clark & Harper, 2000): First, a semi-variogram was constructed to establish the predictability of values from place to place in the study area. The semi-variogram modelled the difference between a value at one location and the value at another according to the distance and between them. Secondly, "kriging" was used to estimate values at unsampled locations. The basic technique of ordinary kriging that we used here used a weighted average of neighbouring samples to estimate the value at an unsampled location. Weights were optimised using the semi-variogram model, given the distance and directional relationships between sampled and unsampled locations. We used the ordinary kriging variance as an estimate of error associated with each prediction (Diggle & Ribeiro, 2007). With this approach, we produced a compilation of EC and pH predictions for each species; given each individual prediction corresponds to an estimate for each location at which the species is recorded. This gives a range of predicted EC and pH values for each species, and from this range we recorded the median and upper quartile (UQ) values. Therefore, for each species, we used four measures to describe soil salinity and alkalinity across its distribution: two describing EC (median and UQ values) and two describing pH (again, median and UQ values). Median values provide species' central tendency with respect to environmental conditions (EC and pH) in their distributions, while UQ values represent more extreme salinity and alkalinity conditions that species encounter within their geographic ranges.

Evaluating prediction of halophytes

An ideal way to evaluate how well the geochemical modelling approach performed in predicting species' salinity and alkalinity tolerance, would be to test species' tolerances experimentally, as well as to take EC and pH measurements at localities where species occur naturally, covering each species range, and then compare those measurements to

our predictions. However, to generate this data, even for one single species, would require considerable amount of time and effort. An alternative way to evaluate the performance of the geochemical modelling is using data that is already available. Although we do not have prior knowledge of alkaline-tolerant species, we have lists of halophytes. These lists might be incomplete, but they are likely to be accurate in the species that are included, as they are based on expert judgment and experimental data. Because halophytes are able to grow in conditions of high salinity, the predicted EC for taxa known to be halophytes should be higher than that for non-halophytes.
Here, we asked whether known halophytes have been reported to occur at higher predicted EC than non salt-tolerant species. First, we extracted the species names of known Australian grass halophytes from a recent study (Bennett et al., 2013), which identified 141 Australian grasses as halophytes (Supplementary Table 1). Then, we applied a parametric Welch two sample t-test to test if predicted EC values (median and UQ) of known halophytes were significantly higher than the rest of the species in our dataset.

Further, we performed the same analysis (Welch two sample t-test) at the genus level, to ask if genera containing halophytes occur in conditions of high predicted EC. There are 234 Australian grass genera in total, 71 of which include at least one known halophyte. We calculated median and upper quartile soil EC values for each genus, based on the observations for all species within that genus. Further, we used a phylogeny-corrected two sample t-test. We estimated the phylogenetic correlation matrix among genera using two phylogenies. One is a well-sampled genus-level topology of Poaceae that includes over 800 genera (Bouchenak-Khelladi et al., 2010). This tree included 226 of the 234 Australian genera and 70 of the 71 genera with known haplotypes. We computed the branch lengths of the topology using the method by Grafen (1989), which gives each node on the tree a ‘height’, corresponding to the number of leaves of the subtree minus one. Each height was scaled so that root height is 1, and then raised at power “rho” (Grafen, 1989). Branch lengths were then calculated as the difference between height of lower and upper nodes. The other phylogeny was a smaller, time-calibrated molecular phylogenetic tree with 298 out of approximately 800 genera of Poaceae (Bouchenak-Khelladi et al., 2010). This tree included 146 of 234 Australian genera and 56 of 71 Australian genera with known haplotypes. We performed the analysis using this tree because, although taxon sampling was limited, it was time-calibrated, and we wanted to ensure that the absence of branch lengths in the larger

phylogenetic tree did not affect our results. We accounted for phylogenetic relatedness in a two sample t-test using Generalized Least Squares (GLS) approach. GLS is a generalised approach for estimating parameters in a linear regression model where observations are not homoscedastic or independent from each other (Martins & Hansen, 1997). The phylogenetic relatedness was accounted for by correcting the covariance matrix among observations according to their phylogenetic relatedness (Martins & Hansen, 1997).

The parametric test compared predicted EC values for halophytic taxa to predicted EC values of the rest of the taxa, and evaluated whether halophytic taxa had higher predicted EC than non-halophytic taxa. Because salt tolerance is not randomly distributed in the grass phylogeny (Bennett et al., 2013), by accounting for phylogenetic relatedness, the phylogenetic test ensured that if a relationship was recovered, it was beyond that expected from phylogeny.

Testing the correlation between salt and alkalinity tolerance

Correlation of taxa occurring in high predicted salinity and alkalinity

We asked whether the taxa found in conditions of high predicted salinity also tended to be found in conditions of high predicted alkalinity. Similar to the previous section, we first calculated the median and UQ EC and pH values for each taxon. We performed this analysis at the species level, testing the correlation between species' median or UQ EC values and species' median or UQ pH values, using the parametric Pearson's product-moment correlation. The same analysis was performed at the genus level, along with a phylogenetic reduced major axis (RMA) regression (Ives et al., 2007), using the two phylogenies described above to estimate the phylogenetic correlation matrix. RMA regression is a type II regression that does not assume causal directionality between values of salinity and alkalinity. The phylogenetic relatedness is accounted for by a similar approach as in GLS (Ives et al., 2007; Martins & Hansen, 1997). Although the parametric test evaluates the correlation between predicted EC and pH for taxa, the phylogenetic test evaluates whether this correlation is because of covariation due to shared ancestry among taxa.

Geographical correlation of salinity and alkalinity

We wanted to tease apart whether any association between predicted salinity and alkalinity values was due to geographical correlation between soil EC and pH. First, to assess the degree to which salinity and alkalinity overlapped on the landscape in areas where Australian grasses are found, we fitted a linear model between predicted values of EC and pH for all occurrence points where Australian grasses were reported. If at localities where predicted EC was high, predicted pH was also high (and vice versa), then species exposed to high salinity were also exposed to high alkalinity (and vice versa).

Second, we tested for the correlation between predicted salinity and alkalinity only for known halophytes, using a parametric Pearson's product-moment correlation. We also tested this relationship at the genus level, only for genera that contain known halophytes, with the parametric Pearson's product-moment correlation, and a phylogenetic reduced major axis (RMA) regression (Ives et al., 2007), using the two phylogenies to estimate the phylogenetic correlation matrix. If salt and alkalinity-tolerance were functionally associated but conditions of salinity and alkalinity were not geographically associated, then salt-tolerant taxa could be found in conditions of both low and high alkalinity. Under these conditions, we would expect a weaker correlation between predicted EC and pH values in salt-tolerant than non salt-tolerant taxa. If salt and alkalinity-tolerance were functionally associated and conditions of salinity and alkalinity were geographically associated, we would expect a stronger correlation between predicted EC values and pH values in in salt-tolerant than non salt-tolerant taxa.

All statistical analyses used log-transformed EC values for normality and were implemented in R (R Core Team, 2014), with Grafen's computation of branch lengths (Grafen, 1989) using the 'compute.brlen' function in 'ape' package (Paradis et al., 2004), the phylogeny-corrected t-test using the 'gls' function in 'nlme' package (Pinheiro et al., 2014), and the phylogenetic RMA regression using the 'phyl.RMA' function in 'phytools' package (Revell, 2012).

Results

Predicting salt and alkalinity tolerance from species distribution modelling

Predicting species salt and alkalinity tolerance from occurrence data

Predicted soil EC for all occurrence points where Australian grasses are found ranged between 0.01 and 10.53 dS m^{-1} and predicted pH ranged from 4.87 to 9.05. The average standard error (as estimated with kriging variance) for predictions across all reported localities was 2.06 dS m^{-1} for EC and 0.93 for pH.

Evaluating prediction of halophytes

Our results (Table 1) show that halophytic species are not found in significantly higher predicted salinity than non salt-tolerant species. However, both analyses (parametric and phylogeny-corrected) at the genus level, considering both median and UQ predicted EC, suggest that genera with known halophytes are found in significantly higher predicted soil EC than genera that do not include known halophytes. Although significantly positive, the absolute difference in EC values between genera with and without known halophytes is small. The predicted EC values for genera with known halophytes only explains about 5% variation of the total of EC values in our dataset (R^2 in Table 1).

Testing the correlation between salt and alkalinity tolerance

Correlation of taxa occurring in high predicted salinity and alkalinity

Our results indicate that species found in conditions of high predicted salinity also tend to be found in conditions of high predicted alkalinity. This is true when considering species' median and UQ EC and pH (Table 1). The same result is found at the genus level, including when accounting for phylogenetic relatedness (Table 2).

Geographical correlation of salinity and alkalinity

The Pearson correlation coefficient (r) can range between -1 (total negative correlation) and 1 (total positive correlation), with 0 denoting no correlation. The value we recovered for the correlation between predicted EC and pH at localities where species were found is very close to 0 (0.0003), suggesting this correlation is extremely weak. Although we found a significant effect ($p < 0.001$), this could be due to a weak relationship in a large amount of data points (N = 354,913). We found a stronger correlation between predicted EC values and pH values for salt-tolerant than for non salt-tolerant taxa, both at the species and genus level (Table 1, Table 2, Figure 1).

Discussion

Predicting salt and alkalinity tolerance from species distribution modelling

The motivation for this study was to explore a possible correlation between salt and alkaline tolerance (Bromham et al., 2013; Bui, 2013; Bui et al., 2014), using Australian grasses as an example. We used a geochemical modelling approach to predict the conditions of salinity and alkalinity in which species occur in their natural distributions (Bui et al., 2014). There are some limitations to this approach. First, our EC predictions were based on measurements in dilute (1:5) solutions compared to the salt concentrations that plants would encounter in saline soils. Predicted electrical conductivity (EC) across localities where grass species were found ranged from 0.01 to 10.53 dS m^{-1}, and halophytes are often described as species that complete their life-cycles in soils of 8 dS m^{-1} and above (Aronson, 1989). Very few localities in our dataset were found above that threshold and only four known halophytes are found in these localities. Nevertheless, our geochemical modelling approach was not used to predict species' absolute tolerances, but relative tolerances that can be used in a comparative framework. Second, it is possible that the geochemical modelling does not accurately capture variation in salinity at the scale that is relevant to ecophysiology. Salinity varies on a micro-scale, depending on many factors, such as climate, lithology, topography, and vegetation (Bui, 2013). Plant distributions can be determined by the distribution of

salinity at that scale, but that will not necessarily be picked up by these landscape-level estimates.

Because of these possible restrictions, we wanted to evaluate the relevance of our geochemical predictions to plant salt tolerance. To do so, we compared predicted salinity values for known halophytic taxa to the rest of the taxa in our dataset. Using a parametric Welch two sample t-test, we found that predicted EC for known halophytes is not significantly higher than that for non-halophytes. Nevertheless, when testing this relationship at the genus level, we found that genera with known halophytes have significantly higher predicted soil EC than genera that do not include known halophytes, using a parametric and a phylogeny-corrected approach. This is likely due to the fact that the list of known halophytes in Australian grasses is much more incomplete than the list of genera with known halophytes. Treating unrecognised halophytes that have high predicted EC values as non-halophytes could contribute to the smaller effect size (R^2 value) in the species-level analyses compared to the genus-level ones, as we show in Tables 1 and 2. We explored two different values to represent predicted EC for each taxon: median and UQ. Our results show that UQ, representing the more extreme values of EC, is better at predicting clades with halophytes, because the effect size (R) is always larger for UQ values than for median values (Table 1, Table 2). It is problematic that some known halophytes are not found in high predicted EC (Figure 1), suggesting that our geochemical approach does not identify salt tolerance successfully. However, our predicted EC values have the potential to identify groups of possible halophytes. The main goal of this study was to investigate the correlation between salt and alkaline tolerance. Therefore, as mentioned above, we aimed at generating relative – rather than absolute - tolerances that can be analysed comparatively for all taxa in the dataset.

Patterns of correlation between salt and alkalinity tolerance

Previous studies have found correlations between different types of ecophysiological strategies related to environmental stress tolerance, particularly to water use efficiency. For example, salt-tolerant grasses have evolved more frequently in lineages with C_4 photosynthesis, potentially because these lineages can control water loss better than C_3 lineages, giving them an advantage to adapt to arid saline environments (Bromham &

Bennett, 2014). A correlation was found between salt tolerance, succulence and C_4 photosynthesis in chenopods (Kadereit et al., 2012), and a similar evolutionary correlation has been found between CAM photosynthesis and succulence (Ogburn & Edwards, 2010), as well as for occupation of bare environments and to adaptation to harsh elemental soils in the Brassicaceae (Cacho & Strauss, 2014).

Our results suggest that salt and alkaline tolerance are associated: we found that species found in conditions of high predicted salinity tend to be found in conditions of high predicted alkalinity (Table 1). This relationship was also recovered at the genus level, including when correcting for phylogenetic relationships (Table 2). This is in agreement with the recent finding that salt and alkaline tolerance are also linked on the phylogeny of Australian *Acacia* (Bui et al., 2014). One possible explanation for the association between taxa in high predicted salinity and alkalinity is the presence of "enablers" in some lineages that can facilitate the evolution of multiple stress resistance within those lineages (Edwards & Donoghue, 2013). It could be that some lineages have traits that provide "stepping stones" to developing both salt and alkaline tolerance: that is, lineages may have traits that do not in themselves confer salt tolerance but make it easier for those lineages to evolve tolerance of saline or alkaline conditions.

However, the correlation we find could also be driven by the overlap of salinity and alkalinity in the landscape (Rengasamy, 2010). We assessed the degree to which predicted salinity and alkalinity correlated in localities where Australian grasses are reported. The correlation between EC and pH at species' localities is significant, but it does not explain much of the variation in our data. Therefore, species exposed to high predicted EC are not necessarily also exposed to high predicted pH at the same localities. For example, as shown in Supplementary Figure 1, the highest predicted EC values are found in both predicted alkaline and acidic soils, and the localities with the highest predicted pH values have low to relatively high predicted EC. Nonetheless, predicted EC values and pH values are more strongly associated for salt-tolerant than for non salt-tolerant taxa (Table 1, Table 2, Figure 1). Therefore, we cannot discount the effect of the overlap of salinity and alkalinity in the landscape in shaping the pattern of correlation we found here. Further research is needed to evaluate how much this overlap contributes to the recovered pattern.

As we have demonstrated, geochemical modelling predictions may provide useful starting points for further investigations of macroevolutionary patterns between salt and alkaline tolerance. However, we did not investigate soil ion chemistry across localities where Australian grasses are found, which affects the correlation between salinity and alkalinity in the soil. For example, soil pH between 7 and 10 mainly reflects anions in solution and when neutral salts such as NaCl or Na_2SO_4 are in solution, sulphate and chloride anions dominate and pH is between 6 and 8 (Rengasamy, 2010). When bicarbonate and carbonate ions dominate, pH rises above 8. At pH above 9, carbonate ions are dominant. Alkaline soils without salt can have pH above 9 but when NaCl is present, pH is lower. However for pH measured in 1:5 soil:water as estimated here, the decrease in pH associated with the presence of NaCl will be diminished by dilution. Also, not all alkaline soils are toxic for plants. For instance, although calcareous soils - abundant in Australia - can be an edaphic barriers to plant radiation [e.g. Nullarbor Plain (Crisp & Cook, 2007)], no toxicity has been observed in lime (CaCO3) dominant soils: although they are alkaline, the solubility of $CaCO_3$ is low and the carbonate concentration is usually around 1 mmolc/L. Given this, we believe more phylogenetic analyses incorporating more complete soil chemistry, as well as testing soil toxicity across sites (Cacho & Strauss, 2014), can lead to more detailed explanations of our reported correlations between salinity and alkalinity for grasses. Further, future investigations could focus on specific traits that might be shared between salinity and alkalinity-tolerances. For example, similar osmotic responses to salinity and alkalinity (Chen et al., 2011; Liu et al., 2010) suggest that some shared mechanisms might be involved in managing water use efficiency under salt and alkaline tolerance. These mechanisms can be investigated experimentally, but a comparative phylogenetic framework may also be useful. For example, species' geochemical predictions can be analysed in a comparative framework that can reveal the degree to which phylogenetic relatedness or spatial autocorrelation can explain the variation in these datasets (Freckleton & Jetz, 2009).

Conclusions

In this study, we used distribution data for Australian grasses combined with geochemical modelling to predict the range of values of soil salinity and alkalinity to

which species are exposed. The aim of this study was to evaluate the use of geochemical modelling in identifying taxa that can tolerate conditions of high salinity and alkalinity. Therefore, our approach was not used to predict species' absolute tolerances, but relative tolerances that can be used in a comparative framework. We find that our geochemical predictions, despite their limitations, can identify known halophytic taxa as present in conditions of relatively high salinity. We also found that grass taxa found in areas of high predicted salinity also tend to be found in conditions of high predicted alkalinity. This pattern could suggest a correlation between salt and alkalinity-tolerance, for example due to the presence of enabling traits that promote the evolution of salinity and alkalinity tolerance. Our approach provides a valuable test of the use of geochemical modelling to predicting abiotic stress tolerances, beyond those related to temperature and precipitation. Further investigations could consider the phylogenetic distribution of specific traits involved in these ecophysiological strategies, ideally by incorporating more comprehensive and finer scale information on variation of geochemistry in the landscape.

Acknowledgments

We thank Dan Warren for helpful comments during this study.

Funding information

This work was supported by the Australian Research Council.

Literature cited

Ackerly, D. D., Dudley, S. A., Sultan, S. E., Schmitt, J., Coleman, J. S., Linder, C. R., Sandquist, D. R., Geber, M. A., Evans, A. S., Dawson, T. E. & Lechowicz, M. J. (2000) The evolution of plant ecophysiological traits: recent advances and future directions. *BioScience 50*, 979-995.
Atlas of Living Australia website at http://www.ala.org.au. Accessed 15 November 2013 [database on the Internet].
Aronson, J. (1989) *HALOPH a data base of salt tolerant plants of the world*. Arizona: Office of Arid Land Studies, University of Arizona.
Ashraf, M. & Foolad, M. R. (2013) Crop breeding for salt tolerance in the era of molecular markers and marker-assisted selection. *Plant Breeding 132*, 10-20.
Bennett, T. H., Flowers, T. J. & Bromham, L. (2013) Repeated evolution of salt-tolerance in grasses. *Biology Letters 9*.
Bernstein, L. (1975) Effects of salinity and sodicity on plant growth. *Annual Review of Phytopathology 13*, 295-312.
Bouchenak-Khelladi, Y., Verboom, G. A., Savolainen, V. & Hodkinson, T. R. (2010) Biogeography of the grasses (Poaceae): a phylogenetic approach to reveal evolutionary history in geographical space and geological time. *Botanical Journal of the Linnean Society 162*, 543-557.
Bromham, L. & Bennett, T. H. (2014) Salt tolerance evolves more frequently in C4 grass lineages. *Journal of Evolutionary Biology 27*, 653-659.
Bromham, L., Saslis-Lagoudakis, C. H., Bennett, T. H. & Flowers, T. J. (2013) Soil alkalinity and salt tolerance: adapting to multiple stresses. *Biology Letters 9*.
Bui, E. N. (2013) Soil salinity: A neglected factor in plant ecology and biogeography. *Journal of Arid Environments 92*, 14-25.
Bui, E. N. (2013) Possible role of soil alkalinity in plant breeding for salttolerance. *Biology Letters 9*.
Bui, E. N., Thornhill, A. & Miller, J. T. (2014) Salt- and alkaline-tolerance are linked in Acacia. *Biology Letters 10*.
Cacho, N. I. & Strauss, S. Y. (2014) Occupation of bare habitats, an evolutionary precursor to soil specialization in plants. *Proceedings of the National Academy of Sciences 111*, 15132-15137.
Chapin, F. S., Autumn, K. & Pugnaire, F. (1993) Evolution of suites of traits in response to environmental stress. *The American Naturalist 142*, S78-S92.
Chen, W., Feng, C., Guo, W., Shi, D. & Yang, C. (2011) Comparative effects of osmotic-, salt- and alkali stress on growth, photosynthesis, and osmotic adjustment of cotton plants. *Photosynthetica 49*, 417-425.
Clark, I. & Harper, W. V. (2000) *Practical geostatistics*. Columbus, OH: Ecosse North American LLC.
Costa, K. C., Navarro, J. B., Shock, E. L., Zhang, C. L., Soukup, D. & Hedlund, B. P. (2009) Microbiology and geochemistry of great boiling and mud hot springs in the United States Great Basin. *Extremophiles 13*, 447-459.
Crisp, M. D. & Cook, L. G. (2007) A congruent molecular signature of vicariance across multiple plant lineages. *Molecular Phylogenetics and Evolution 43*, 1106-1117.
de Caritat, P. & Cooper, M. (2011) *National Geochemical Survey of Australia: The Geochemical Atlas of Australia. Record 2011/20*. Canberra: Geoscience Australia.
Deinlein, U., Stephan, A. B., Horie, T., Luo, W., Xu, G. & Schroeder, J. I. (2014) Plant salt-tolerance mechanisms. *Trends in Plant Science*.
Diggle, P. & Ribeiro, P. (2007) *Model-based geostatistics*: Springer.

Edwards, E. J. & Donoghue, M. J. (2013) Is it easy to move and easy to evolve? Evolutionary accessibility and adaptation. *Journal of Experimental Botany*.
Farrell, R., Bell, D., Akilan, K. & Marshall, J. (1996) Morphological and physiological comparisons of clonal lines of *Eucalyptus camaldulensis*. II. Responses to waterlogging/salinity and alkalinity. *Functional Plant Biology 23*, 509-518.
Flowers, T. J. & Colmer, T. D. (2008) Salinity tolerance in halophytes. *New Phytologist 179*, 945-963.
Flowers, T. J., Galal, H. K. & Bromham, L. (2010) Evolution of halophytes: multiple origins of salt tolerance in land plants. *Functional Plant Biology 37*, 604-612.
Flowers, T. J., Troke, P. F. & Yeo, A. R. (1977) The mechanism of salt tolerance in halophytes. *Annual Review of Plant Physiology 28*, 89-121.
Freckleton, R. P. & Jetz, W. (2009) Space versus phylogeny: disentangling phylogenetic and spatial signals in comparative data. *Proceedings of the Royal Society B: Biological Sciences 276*, 21-30.
García, I. & Mendoza, R. (2014) *Lotus tenuis* seedlings subjected to drought or waterlogging in a saline sodic soil. *Environmental and Experimental Botany 98*, 47-55.
Grafen, A. (1989) The phylogenetic regression. *Philosophical Transactions of the Royal Society of London B 326*, 119-157.
Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters 8*, 993-1009.
Islam, A. K. M. S., Edwards, D. G. & Asher, C. J. (1980) pH optima for crop growth. *Plant and Soil 54*, 339-357.
Ives, A. R., Midford, P. E. & Garland, T., Jr. (2007) Within-species variation and measurement error in phylogenetic comparative methods. *Systematic Biology 56*, 252-270.
Kadereit, G., Ackerly, D. & Pirie, M. D. (2012) A broader model for C4 photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae s.s.). *Proceedings of the Royal Society B: Biological Sciences 279*, 3304-3311.
Kozak, K. H., Graham, C. H. & Wiens, J. J. (2008) Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology & Evolution 23*, 141-148.
Liu, J., Guo, W. Q. & Shi, D. C. (2010) Seed germination, seedling survival, and physiological response of sunflowers under saline and alkaline conditions. *Photosynthetica 48*, 278-286.
Macur, R. E., Langner, H. W., Kocar, B. D. & Inskeep, W. P. (2004) Linking geochemical processes with microbial community analysis: successional dynamics in an arsenic-rich, acid-sulphate-chloride geothermal spring. *Geobiology 2*, 163-177.
Martins, E. P. & Hansen, T. F. (1997) Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *The American Naturalist 149*, 646-667.
Munns, R. (2005) Genes and salt tolerance: bringing them together. *New Phytologist 167*, 645-663.
Munns, R. & Tester, M. (2008) Mechanisms of salinity tolerance. *Annual Review of Plant Biology 59*, 651-681.
Niinemets, Ü. & Valladares, F. (2006) Tolerance to shade, drought, and waterlogging of temparate Northern Hemisphere trees and shrubs. *Ecological Monographs 76*, 521-547.
Nishiuchi, S., Fujihara, K., Liu, S. & Takano, T. (2010) Analysis of expressed sequence tags from a NaHCO3-treated alkali-tolerant plant, *Chloris virgata*. *Plant Physiology and Biochemistry 48*, 247-255.
Ogburn, R. M. & Edwards, E. J. (2010) The ecological water-use strategies of succulent plants. *Advances in Botanical Research 55*, 179-225.

Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics 20*, 289-290.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-117. 2014.
R Core Team. R: A Language and Environment for Statistical Computing. 2014.
Rengasamy, P. (2010) Soil processes affecting crop production in salt-affected soils. *Functional Plant Biology 37*, 613-620.
Revell, L. J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution 3*, 217-223.
Reysenbach, A.-L. & Shock, E. (2002) Merging Genomes with Geochemistry in Hydrothermal Ecosystems. *Science 296*, 1077-1082.
Rozema, J. & Flowers, T. (2008) Crops for a salinized world. *Science 322*, 1478-1480.
Rozema, J. & Schat, H. (2013) Salt tolerance of halophytes, research questions reviewed in the perspective of saline agriculture. *Environmental and Experimental Botany 92*, 83-95.
Saslis-Lagoudakis, C. H., Moray, C. & Bromham, L. (2014) Evolution of salt tolerance in angiosperms: A phylogenetic approach. In: Rajakaruna, N, Boyd, B, Harris, T (Eds.), *Plant Ecology and Evolution in Harsh Environments*, (pp. 77-95)
Shabala, S. (2013) Learning from halophytes: physiological basis and strategies to improve abiotic stress tolerance in crops. *Annals of Botany 112*, 1209-1221.
Shavrukov, Y. (2012) Salt stress or salt shock: which genes are we studying? *Journal of Experimental Botany 64*, 119-127.
Tuteja, N. (2007) Mechanisms of high salinity tolerance in plants., *Osmosensing and osmosignaling*, (pp. 419-438). San Diego, CA: Elsevier Academic Press.
United States Salinity Laboratory Staff. (1969) *Diagnosis and improvement of saline and alkali soils*. Washington, DC: United States Department of Agriculture.
Warren, D. L., Glor, R. E. & Turelli, M. (2008) Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution 62*, 2868-2883.
Yang, C., Chong, J., Li, C., Kim, C., Shi, D. & Wang, D. (2007) Osmotic adjustment and ion balance traits of an alkali resistant halophyte *Kochia sieversiana* during adaptation to salt and alkali conditions. *Plant and Soil 294*, 263-276.
Yang, C., Shi, D. & Wang, D. (2008) Comparative effects of salt and alkali stresses on growth, osmotic adjustment and ionic balance of an alkali-resistant halophyte *Suaeda glauca* (Bge.). *Plant Growth Regulation 56*, 179-190.

Table 1: Results of tests for the comparison of predicted EC values for known halophytes vs. non salt-tolerant species, and for the correlation between salinity and alkalinity conditions in Australian grass species. Alternative hypotheses are listed in the first column. The variable tested (median or UQ) for species' salinity and/or alkalinity is given in the second column. Each hypothesis was tested with parametric test. *t*-statistic and R^2 values are reported for each test. Statistics significant at 0.05 level are marked with an asterisk; significant at 0.005 level are marked with a double asterisk. Significant statistics support the alternative hypotheses.

Alternative hypothesis	Variable	Parametric
Known halophytes are found in conditions of higher predicted salinity than non salt-tolerant species	Median	t_{185}=0.54 R^2=0.00
	Upper quartile	t_{185}=1.14 R^2=0.01
Species found in conditions of high predicted salinity also tend to be found in conditions of high predicted alkalinity	Median	T_{1385}=29.63** R^2=0.39
	Upper quartile	T_{1385}=35.96** R^2=0.48
Known halophytes found in conditions of high predicted salinity tend to be found in conditions of high predicted alkalinity	Median	T_{139}=12.33** R^2=0.52
	Upper quartile	T_{139}=17.88** R^2=0.70

Table 2: Results of tests for the comparison of predicted EC values for genera including known halophytes vs. those not including halophytes, and for the correlation between salinity and alkalinity conditions in Australian grass genera. Alternative hypotheses are listed in the first column. The variable tested (median or UQ) for salinity and/or alkalinity of a given taxon is given in the second column. Tests for each hypothesis include a parametric and two phylogeny-corrected analyses. The phylogeny-corrected analyses were performed on a complete genus-level phylogenetic tree of grasses (Complete column) and a smaller, time-calibrated phylogenetic tree (Calibrated column) from a previous study (Bouchenak-Khelladi et al., 2010). *t*-statistic and R^2 values are reported for each test. Statistics significant at 0.05 level are marked with an asterisk; significant at 0.005 level are marked with a double asterisk. Significant statistics support the alternative hypotheses.

Alternative hypothesis	Variable	Parametric	Phylogeny-corrected	
			Complete	Calibrated
Genera with known halophytes are found in conditions of higher predicted salinity than genera without known halophytes	Median	t_{186}=3.25** R^2=0.04	t_{224}=3.03** R^2=0.04	t_{144}=2.45* R^2=0.04
	Upper quartile	t_{209}=3.89** R^2=0.06	t_{144}=4.44** R^2=0.08	t_{144}=2.46* R^2=0.04
Genera found in conditions of high predicted salinity tend to be found in conditions of high predicted alkalinity	Median	t_{232}=11.18** R^2=0.35	t_{198}=12.60** R^2=0.28	t_{116}=9.02** R^2=0.53
	Upper quartile	t_{232}=15.60** R^2=0.51	t_{144}=16.68** R^2=0.56	t_{109}=13.75** R^2=0.70
Genera with known halophytes found in conditions of high predicted salinity tend to be found in conditions of high predicted alkalinity	Median	t_{69}=6.71** R^2=0.40	t_{55}=3.96** R^2=0.35	t_{45}=6.22** R^2=0.62
	Upper quartile	t_{69}=9.05** R^2=0.54	t_{55}=5.96** R^2=0.56	t_{43}=10.33** R^2=0.72

Figure 1: Correlation between predicted soil salinity and alkalinity for Australian grass genera. The predicted salinity and alkalinity of a given genus is measured as the median (A) and upper quartile (B) value of all predictions of electrical conductivity (EC) or pH, respectively, for all localities where species of that genus occur in mainland Australia. Black dots are genera that do not include any known halophytes and red dots are genera that include known halophytes.