

Polyploid clades tend to be more speciose than their putative sister clades

Jack Oyston (✉ jwo22@bath.ac.uk)

The University of Bath <https://orcid.org/0000-0002-8291-7271>

Mark Wilkinson

The Natural History Museum

Marcello Ruta

University of Lincoln <https://orcid.org/0000-0002-6151-0704>

Matthew Wills

University of Bath

Article

Keywords:

Posted Date: February 21st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1317290/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Striking asymmetries of species richness characterise the tree of life, and understanding their causes is a major research agenda in evolutionary biology. Genome duplications have been proposed as one mechanism by which lineages acquire greater potential for rapid genetic evolution following the relaxation of pleiotropic constraints. This enhanced evolvability might facilitate faster rates of morphological change and speciation. We provide a comprehensive test of the empirical association between polyploidy – a marker of genome duplication – and species richness in 356 pairs of extant polyploid and non-polyploid sister clades. Comparing sister clades controls for divergence time and many ecological and morphological properties of groups. Clades with inferred basal genome duplications have significantly higher species diversity than their sister clades, both for the whole dataset, major groups of animals and plants within it and at the genus and family levels. Genome duplications therefore offer one possible, widespread cause of heterogeneities in species richness.

Introduction

Biodiversity is unevenly distributed across the major branches of the tree of life. Even groups that diverged from the same common ancestor – so called “sister groups” – often have strikingly contrasting levels of species richness. One of the most extreme examples of these imbalances in biodiversity can be found in arthropods. While there are an estimated 5.5 million species of insects¹ there are just 28 species² of their nearest extant relatives, the remipede crustaceans^{3–5}. These remarkable differences in species diversity may reflect differences in bodyplan, habitat and ecology^{6–8}. For example, insects have well-defined tagmosis with three pairs of strongly differentiated limbs, while remipedes have between ten and thirty-two pairs of homonomous trunk limbs. Insects live in virtually all terrestrial and many freshwater habitats, while remipedes are restricted to anchialine cave systems⁹. Insects occupy numerous guilds and display a bewildering array of trophic strategies and adaptations, while remipedes are raptorial predators. Many of the differences in morphology, ecology and habitat that might explain differences in species richness are often unique to particular clades and particular cases. While there may also be environmental and geographical factors that link differences in biodiversity to environmental and geographic factors, these differences could be underpinned by genetic factors that strongly impact speciation rate. Here we explore the potential for genome duplications to facilitate species diversification as one such generality.

Gene duplication is the most important source of novel genes, yielding an estimated 50% of prokaryote¹⁰ and 90% of eukaryote genes¹¹. Studies of the human genome suggest that duplications have continuously generated paralogues since the origin of chordates¹², and are vital for the evolution of gene regulatory networks¹³. Small-scale duplications of genes (SSDs) may underpin increases in gene family sizes, which itself has numerous macroevolutionary correlates, including increasing encephalisation and expansion of the neocortex in mammals^{14,15}. However, there are few studies^{14,15} assessing the extent and impact of SSDs directly¹⁶.

Most studies correlating genome duplication to macroevolutionary patterns and diversity have focused on whole genome duplications (WGDs), generating polyploid taxa with more than two complete sets of chromosomes^{17,18}. For example, paleopolyploidy events (ancient WGDs close to or at the root of major clades¹⁹) may have facilitated key innovations and subsequent radiations in angiosperms, particularly rosids and grasses²⁰. Similar events occurred early in the evolution of other groups of seed plants²¹, as well as in fungi such as baker's yeast *Saccharomyces*^{22,23}. Examples of major paleopolyploid animal clades are more controversial. At least two rounds of genome duplication occurred early in the evolution of vertebrates; one at the root of vertebrates and another close to the root of gnathostomes. These appear to correlate with key innovations in the vertebrate bodyplan²⁴, although fossils complicate this pattern²⁵. Ancient WGDs are difficult to detect after duplicated genes have either diverged and acquired novel functions, or been edited out, and most of our understanding of polyploidy therefore derives from relatively recent events.

Ancient WGDs are difficult to detect after duplicated genes have either diverged and acquired novel functions, or been lost through natural selection. Most of our understanding of polyploidy therefore derives from relatively recent events. Polyploidy is widespread in living plants, particularly angiosperms²⁶ and ferns²⁷, with many groups showing successive rounds of genome duplication in their histories²⁸. Perhaps 15% of angiosperm and 31% of fern speciation is directly linked to a ploidy increase²⁹⁻³¹. Within the monocotyledons, grasses are highly polyploid³² and have informed studies of model polyploid genomes^{33,34}. Polyploidy is also common in 'true' dicotyledon groups such as the Brassicaceae³⁵, Fabaceae³⁶, Violaceae³⁷ and Orchidaceae³⁸. Although thought to be rare in most animals³⁰, polyploidy is being identified in an increasing number of amphibians^{39,40} and fish^{41,42}. Polyploidy is well documented in teleosts, especially salmonids⁴³, catfish⁴⁴ and carp⁴⁵. Invertebrate examples include insects⁴⁶, crustaceans⁴⁷ and molluscs⁴⁸.

Duplicate genes in polyploids appear to benefit from relaxed selection pressures and may persist in the genome⁴⁹, thereby accumulating mutations and evolving novel functions 'under the radar' of normal selection pressure⁵⁰. Through this process of neofunctionalization, polyploidy is thought to be the most important source of novel genes and their functions⁵¹, greatly contributing to the expansion of gene families^{52,53}. Gene duplication through polyploidy followed by modification of the resulting paralogues may offer a mechanism for evolving the regulatory networks underpinning aspects of anatomical patterning, such as floral structure in angiosperms^{54,55} and limb arrangement in arthropods⁵⁶. If this is true then polyploid organisms should, all else being equal, show evidence of greater 'evolvability' or adaptive variation than closely related groups with a similar ecology and morphology⁵⁷, and might therefore be expected to give rise to clades with higher diversity. The hypothesis that genome duplication plays an important and widespread role in diversifications on a macroevolutionary scale across the tree of life has not been tested until now. In this study, we compile documented instances of polyploidy for macroscopic eukaryotes to identify pairs of polyploid and non-polyploid sister groups. Using the most recent and accurate published diversity estimates available, we then tested whether polyploid clades have greater species richness than their non-polyploid sister clades.

Results

Unsurprisingly given the broad taxonomic coverage of our dataset (356 clade pairs, 153 animals, 203 plants), species richness varied markedly in both polyploid and non-polyploid clades (Fig. 1A), from branches comprising a single species to hyper-speciose clades (82,320 species in true weevils). Polyploid clades were, on average, significantly more speciose than their non-polyploid sister clades (Table 1: Wilcoxon $V = 45769$, $p\text{-value} = 6.734 \times 10^{-16}$; polyploid: mean = 406 species, median 52 species; non-polyploid: mean = 93 species, median = 15 species) and contained on average 79.5% of the species within each clade pair (Fig. 1B). While the distribution of species numbers in polyploid clades was right skewed by a few clades with extremely high species diversity, the significance of the difference was impervious to the removal of these outliers (polyploid: mean = 123 species, median = 48 species; non-polyploid: mean = 49, median = 14; $V = 42484$, $p\text{-value} = < 2.2 \times 10^{-16}$). As some of the pairs had at least one genus containing only a single species (35 of the 356 pairs), we wanted to ensure that the greater species richness of polyploid clades was not the result of monotypic non-polyploid genera. We find the result holds when clade pairs with monotypic genera are omitted (Wilcoxon $V = 36360$, $p\text{-value} = 1.312 \times 10^{-12}$).

Similar patterns were also found in animal and plant clades analysed separately, as well as in major animal and plant subclades (Fig. 2). The difference in animals was especially striking (Fig. 2A: Wilcoxon signed-rank $V = 8825$, $p\text{-value} = 9.807 \times 10^{-09}$) as polyploid clades contained, on average, almost 75% of the species within each clade pair. Both vertebrates (Fig. 2B: Wilcoxon $V = 3025$ $p\text{-value} = 2.895 \times 10^{-05}$) and invertebrates (Fig. 2C: Wilcoxon $V = 1542$, $p\text{-value} = 7.445 \times 10^{-05}$) analysed individually showed patterns like those found in animals as a whole (Supplementary Fig. 1). Significant differences in species richness were also recovered for the 'fish' (non-tetrapod vertebrates), insect, crustacean and mollusc datasets (Table 1). In both tetrapods as a whole and lissamphibians specifically (Supplementary Fig. 2), polyploid clades tended to be more speciose, but the presence of a small number of highly speciose non-polyploid clades rendered Wilcoxon signed-rank tests of species richness non-significant. Sign tests carried out on both tetrapods ($S = 32$, $p\text{-value} = 0.0293$) and lissamphibians ($S = 20$, $p\text{-value} = 0.0357$) did show significantly higher species richness in polyploid clades, as these statistical tests ignore effect size. Reptiles and annelids showed little evidence of higher species numbers within polyploid clades (Table 1).

Plants also had, on average, significantly more speciose polyploid (mean = 220, median = 70) than non-polyploid (mean = 112, median = 15) sister clades (Fig. 2D: Wilcoxon $V = 14468$, $p\text{-value} = 1.072 \times 10^{-08}$). Like animals, most of the highly speciose outliers were in polyploid clades, although there were a small number of highly diverse non-polyploid clades. Angiosperms (Fig. 2E: Wilcoxon $V = 5501.5$, $p\text{-value} = 5.018 \times 10^{-05}$) and non-angiosperms (Fig. 2F: Wilcoxon $V = 2155$, $p\text{-value} = 3.595 \times 10^{-05}$) analysed individually also showed patterns similar to plants as a whole. Similar patterns were recovered for plant subclades (Supplementary Fig. 3), although these differences were found to be significant only for dictyots (Wilcoxon $V = 2654.5$, $p\text{-value} = 0.003779$) and ferns (Wilcoxon $V = 1852$, $p\text{-value} = 0.0001179$). Polyploid clades were markedly more speciose in all other groups with statistical power being limited by smaller sample sizes.

Although our dataset contained clades at different taxonomic levels, separate analyses of genera and families (Fig. 3) recovered similar patterns to those above (Table 2, Supplementary Fig. 4). There were insufficient clade pairs at other taxonomic levels (2 pairs of orders, 1 pair of suborders, 4 pairs of subfamilies) to test for differences. On average, polyploid genera were found to contain significantly more species than their non-polyploid sister genera (Fig. 3A: polyploid: mean = 129, median = 46; non-polyploid: mean = 70, median = 13; Fig. 3B: Wilcoxon $V = 36829$, $p\text{-value} = 5.645 \times 10^{-14}$). Test results in most subgroups were similar to those in the total combined dataset (Table 3), although both tetrapods (Wilcoxon test: $V = 767.5$, $p\text{-value} = 0.03166$) and lissamphibians (Wilcoxon $V = 285.5$, $p\text{-value} = 0.02105$) returned significant differences in the genus only dataset in contrast to the dataset of clades at all taxonomic levels. While genera constituted the large majority of our sample (321 out of 356 clade pairs), the pattern of higher species richness in polyploid clades holds in the sample of higher taxonomic ranks generally (Wilcoxon $V = 493$, $p\text{-value} = 0.002833$), as well as at the level of families specifically (Fig. 3C: polyploid: mean = 3459, median = 251; non-polyploid: mean = 322, median = 57; Fig. 3D: Wilcoxon $V = 311$, $p\text{-value} = 0.0127$).

Finally, polyploid clades were found to show significantly higher species richness than non-polyploid clades without taking into account the size of the difference, where there were differences in species richness sign tests reported a significantly higher number of polyploid clades which were more speciose than their non-polyploid counterparts (Table 4), in most taxonomic groups for both the whole dataset and for genera specifically.

Discussion

Polyploid clades tend to contain more species than their non-polyploid sister clades

Clades with polyploid taxa that are inferred to have undergone genome duplications at or close to their originations contain, on average, significantly more species than sister clades in which such duplication events did not occur. This striking difference was found across nearly all major clades and in nearly every major subgroup of organisms tested. This strongly implies that polyploidy is an important evolutionary cause of diversification¹⁸. There have been a small number of studies investigating the association between ploidy level and diversity in specific plant clades^{30,58,59}. Although these investigations all identified an association between polyploidy and higher taxonomic diversity, it has been unclear whether this results from enhanced evolvability or higher speciation rates in plant clades or whether the effect is driven by reproductive isolation in the immediate aftermath of the WGD events themselves. The taxonomic generality and strength of the effect observed here suggests that the role of WGD events as a cause of diversity may have been underestimated¹⁹. Rather than being prone to rapid extinction^{60,61}, or leading to evolutionary dead ends^{62,63}, polyploids appear to propagate persistent lineages that subsequently diversify more readily than non-polyploid close relatives.

Polyploidy is a widespread correlate of species richness in plants

Polyploidy is extremely common in living plants, particularly angiosperms²⁶ and ferns²⁷, with many groups showing successive rounds of genome duplication in their history²⁸. Perhaps 15% of angiosperm and 31% of fern speciation is directly linked to a ploidy increase^{29–31}. Within the monocotyledons, grasses are highly polyploid³² and have informed studies of model polyploid genomes^{33,34}. Polyploidy is also common in ‘true’ dicotyledon groups such as the Brassicaceae³⁵, Fabaceae³⁶, Violaceae³⁷ and Orchidaceae³⁸.

Despite a long history of polyploidy research in plants^{64–66}, the prevalence of diploid taxa^{67,68} meant that polyploidy was not widely considered as a possible cause of diversity. This assumption can now be challenged⁶⁹ and studies of polyploidy have received renewed interest⁷⁰. The composition of our dataset suggests that polyploidy is particularly common in derived angiosperms (particularly eudicots) and ferns but much rarer in gymnosperms and more ‘basal’ angiosperms (ANA clade plus magnoliids and Chloranthales). Prior studies have stressed the link between polyploidy and speciation in angiosperms⁷¹, with most major angiosperm clades inferred to have polyploid ancestry⁷². While genomic changes such as WGDs have been found to correlate with the origins of major gymnosperm clades as well as the root of gymnosperms itself, more recent shifts in gymnosperm diversification seem to be decoupled from putative WGDs⁷³. We were able to identify few documented instances of natural polyploids in basal angiosperms and gymnosperms, either because polyploidy is genuinely rare⁷⁴, or because any gene duplications were sufficiently ancient to obfuscate its signal²¹. Polyploidy appears to be extremely common in *Magnolia*⁷⁵ but has not been investigated in related magnoliids and Chloranthales. Hence, while polyploid magnoliids are more species rich than their non-polyploid sister taxon, the sample size (9 clade pairs) is small and the difference is not significant. Despite being common, polyploidy in ferns has only recently received systematic attention^{76,77}, and all of our polyploid fern clades were genera. The absence of clear patterns in ploidy level at higher taxonomic levels could be one reason why fern polyploidy has received relatively little study.

Polyploidy is also a widespread correlate of species richness in animals

Although thought to be rare in most animals³⁰, polyploidy is being identified in an increasing number of amphibians^{39,40} and fishes^{41,42}. It is well documented in teleosts, especially salmonids⁴³, catfish⁴⁴ and carp⁴⁵. Invertebrate examples include insects⁴⁶, crustaceans⁴⁷ and molluscs⁴⁸.

The evolutionary significance of polyploidy in animals is contentious and discussion has largely focused on ancient Whole Genome Duplication events (WGDs). For example, two WGDs have been proposed close to the origin of vertebrates²⁴ and another at the origin of bony fishes⁷⁸ which, it has been suggested, helped facilitate the evolution of novel morphologies and more complex phenotypes⁷⁹. The inclusion of fossil data reduces support for this close temporal association in favour of more protracted evolutionary consequences²⁵. There are very few living basal vertebrates and gnathostomes and these have very long evolutionary branches separating them from other vertebrate groups and from their last common ancestors with other clades, making it difficult to assess the impact of polyploidy in these cases. However, at least two orders of ray-finned fish (Acipenseriformes and Salmoniformes) appear to be entirely polyploid, suggesting that more recent genome duplications have also played an important role in the evolution of teleosts⁴². Polyploid

families are also common within Siluriformes (catfish) and Cypriniformes (carp and minnows), while polyploid genera are found within most teleost groups. Paralogs of developmentally relevant genes often appear to be conserved in teleosts⁸⁰ and at least some of them are associated with physiological traits unique to this group, including pigmentation types and colour patterns⁸¹. Polyploidy seems to be extremely rare in chondrichthyans, possibly because inferred substitution rates are much lower in cartilaginous fish than in other vertebrates⁸², and particularly in comparison with the fast rates in teleosts⁸³.

The large majority of other polyploid vertebrate clades are amphibians, with some polyploids also documented in reptiles. Ploidy levels in these groups appear less evolutionarily conserved than in teleosts, with polyploid and non-polyploid taxa often being closely related. As a result, all but one of the clade pairs analysed were genera, albeit spread broadly across the major groups of lissamphibians and squamates. Although polyploidy has been recognised as common in particular groups of anurans, including the clawed frogs⁸⁴, it has only been linked to speciation in a few cases^{85,86}. There are also documented cases of polyploidy in birds and mammals, although these are far more tentative and controversial. ZZW triploidy is known to occur in the domestic chicken, although embryo mortality is extremely high⁸⁷, as well as in the Blue-and-Yellow Macaw⁸⁸. Within mammals, the mountain vizcacha rat has been characterised as an allotetraploid with a hybrid origin⁸⁹ although this characterisation is questionable⁹⁰.

Polyploidy in vertebrates has been studied in greater detail than in other animals⁹¹. Despite this, we were able to find numerous documented cases of polyploid invertebrates, although just over half of our documented clades were insects (32 out of 62). Despite containing well over 5.5 million species¹, the number of polyploid insect clades is thought to be less than 100⁴⁶. We suspect that this discrepancy may result from a disproportionate focus on insects in the literature, which are appealing for three reasons. Firstly, the segmented bodyplan of arthropods make them the ideal model for studying the role of *Hox* gene changes and duplications in determining the expression of morphological traits such as limb identity^{92,93}. Secondly, the fruit fly *Drosophila* is the most widely used model for studies of genetic evolution and expression in animals. Thirdly, insects have extraordinarily high diversity and are major components of nearly all terrestrial ecosystems despite their relatively conserved bodyplan. The evolution of novel *Hox* gene regions is thought to have facilitated morphological diversification in arthropods generally and insects in particular^{94,95}. While it is possible that polyploidy is associated with diversification of gene regulatory networks, such as *Hox* genes in insects, we cannot rule out the possibility that polyploidy is simply more poorly documented in other invertebrate groups.

While the majority of animal groups in our sample showed significantly higher diversity in polyploid clades, reptiles, and annelid worms did not. However, sample sizes for these two groups were small, and in each case polyploid clades still contained more species on average. We think it likely that our inability to detect differences in these clades reflect their under-representation in the literature documenting polyploidy. Studies of oligochaete annelids have suggested little correlation between genome size and life history traits, with the possible exception of parthenogenesis in highly polyploid earthworms⁹⁶. Polyploidy may be rare in annelids with the exception of some highly polyploid genera⁹⁷ but there are no comprehensive assessments of polyploidy in the phylum. Polyploidy in reptiles appears entirely restricted to the squamates,

although highly labile and variable within that group. Viable squamate polyploids appear to be exclusively triploids⁹¹, likely arising through the mating of diploid hybrids with sexual lineages⁹⁸. There is little evidence for strong ecological⁹⁹, or phenotypic¹⁰⁰ differences between parthenogenetic triploids and sexual diploids in the few squamate taxa where polyploidy has been studied in detail.

Polyploid clades at multiple taxonomic levels are more speciose than their non-polyploid sisters

Clades inferred to have undergone genome duplications close to their origins show significantly higher species richness at multiple taxonomic levels. Separate analyses of genera and higher clades both reveal a similar pattern, as do analyses of families considered alone. This makes it unlikely that our results are strongly influenced by the scale at which clades are sampled, as well as by any non-independence of nested clades.

Documented cases of polyploid groups are far more common at the level of genera (321 out of 356 clade pairs) than at higher taxonomic levels. As closely related taxa are, typically, more similar genetically, it may be unsurprising to observe that polyploidy occurs convergently numerous times within the same clade. Ultimately, polyploidy first arises within small populations. Large clades diversifying after genome duplications will be harder to detect as the signal of the ploidy event is overwritten by the subsequent loss of genes and mutation in the retained genes¹⁰¹. While genome duplications are therefore inevitably most common at lower taxonomic levels, the effect may be exaggerated due to our limited ability to detect such events in large clades.

Our study provides empirical evidence that genome duplications correlate with greater diversification across the tree of life. Most previous attempts to investigate the distribution and variation in polyploidy have documented its occurrence at low taxonomic levels in specific groups of organisms^{102,103}, while consideration of its evolutionary significance at the macroevolutionary scale has focused on ancient whole genome duplications^{20,41}. While this study presents compelling evidence of a more general relationship between polyploidy and diversity, species richness is a simplistic measure of diversity that fails to capture many aspects of ecological and morphological variation. Defining clades with ancestral genome duplication events using the literature on extant polyploids lends itself well to studying a broad taxonomic sample. However, it ignores the potentially complex history of genetic changes in different lineages within clades. Further work relating changes in the genome with rates and amounts of morphological evolution may help to explain the ubiquitous pattern seen in this study.

How does polyploidy facilitate diversification?

Polyploidy is often implicated in the diversification and evolution of key traits and is associated with a number of physiological effects¹⁰⁴. Polyploidy can alter patterns of gene expression which lead to increases in cell size and hence tissue growth¹⁰⁵ as well as altering cell architecture and regulatory mechanisms¹⁰⁶. Direct, simple changes in traits like body size are rare¹⁰⁷, as in most cases developmental mechanisms regulate, eliminate or modify these effects¹⁰⁸. Instead polyploidy often affects gene expression in complex ways to produce a broad range of 'hybrid vigour' effects¹⁰⁹, including downregulation as well as the upregulation of genes¹¹⁰. While polyploidy can introduce errors into cell

division that increase the likelihood of sterility¹¹¹ and other deleterious consequences^{112,113}, certain groups (e.g. plants) have genetic and developmental architecture that mitigate these effects¹¹⁴. In plants, sterility is often countered by a greatly increased capacity for selfing¹¹⁵, allowing many polyploids to reproduce asexually by parthenogenesis¹¹⁶ which can also result in instant reproductive isolation and hence speciation¹¹⁷. However, polyploidy may also cause a failure of self-incompatibility mechanisms which allow traits such as dioecy to evolve¹¹⁸

The developmental trajectories of many groups have become more complex through time¹¹⁹⁻¹²². This is because the addition of new developmental stages is constrained by existing developmental pathways^{123,124} and also because genes, developmental pathways, tissues and organs often acquire a greater number of increasingly integrated functions, and thus pleiotropy, over time¹²⁵. The accumulation of pleiotropic effects makes it increasingly likely that mutations will deleteriously affect at least one process or pathway as genes acquire multiple functions¹²⁶. Given increasing complexity, one might therefore expect the evolutionary flexibility of organisms to decline over time as bodyplan development becomes more canalised^{127,128}. Canalisation may help to explain how closely related organisms sharing the same genetic architecture evolved the same phenotypic traits in parallel¹²⁹⁻¹³¹ and may be relatively common even at the more fundamental genetic level, constraining change along 'lines of least resistance'¹³². Genome duplications may offer a mechanism by which lineages can circumvent the worst effects of pleiotropy¹³³, freeing up copies of genes that can – albeit very occasionally – acquire new functions and thereby increase evolvability¹³⁴⁻¹³⁶.

Methods

Sampling polyploid taxa from the published literature

Biological publications spanning the years 1950 to 2018 were searched for known cases of polyploid taxa, using the names of major clades and “polyploidy” as keywords. Important source publications were Mable et al. 2011 for amphibians and fish⁴¹, Otto and Whitton 2000 for reptiles and invertebrates³⁰, Song et al. 2012 for angiosperms⁹¹ and Wood et al. 2009 for vascular plants³¹. The resulting taxa were organised into nested groups by their taxonomy and sampling effort was focused on obtaining as reasonable sample size for as many groups as possible (Supplementary Table 1). In this way we maximised the number of clades at the rank of class and above represented in our dataset over accurately representing the proportion of polyploid taxa in different groups. As a result, the literature of clades for which polyploids were rarely documented (e.g. annelids) was more intensively sampled than that of clades for which polyploidy was extremely common (e.g. flowering plants). Instances of somatic polyploidy (endopolyploidy) were discounted as these are extremely common even in diploid organisms, where it often occurs only in particular cell types, and so represents a clearly distinct phenomenon than that of whole genome duplication we are seeking to investigate here. Similarly, instances of artificially induced polyploidy were discounted as because these occurrences did not occur naturally there is no reason to suppose they could be directly linked to the evolution or diversification of the clade in which they occurred.

Defining clades with ancestral genome duplications

In most cases the literature clearly defined clades of organisms which are known to be exclusively or ancestrally polyploid, and these were then used to define the polyploid clade resulting from an ancestral genome duplication event (hereafter: polyploid clades). In some cases assignments in the literature were conflicting or more controversial, in which case documented polyploid taxa were compared to close relatives (e.g. species in the same genus, genera in the same family). If polyploidy was found to occur in related taxa, then the least inclusive taxonomic ranking would be used to define the clade containing that taxon and the related taxa as having a basal genome duplication (polyploid). If one or more of the related taxa were not polyploid, the group was classed as non-polyploid. Although this a conservative approach is likely to result in some polyploid groups not being included in the dataset this helped to reduce the risk of overestimating species richness in polyploid clades by mistakenly including non-polyploid taxa. This approach was also used when polyploidy was documented at the sub-generic or species level. Although comparing specific clades of species and subspecies within genera would be desirable to more accurately reflect all polyploidy events, phylogenies at this level often poorly supported with high proportions of missing taxa. This makes identifying sister clades confidently impossible in many cases, even when relationships of taxa are resolved, as such relationships are very likely to change in the near future. We also sought to include extinct taxa in our estimates wherever possible and saw comparisons between genera as the lowest rank at which we could reasonably attempt to utilize fossil data. Similarly, in cases where species showed both diploid and polyploid subspecies or species morphs, the entire species was classed as polyploid. In some groups, such as angiosperms, it is common to have taxa of increasing ploidy level nested within each other (e.g. hexaploids nested within a clade of tetraploids). In all cases where it was possible to delimit ploidy level to nested monophyletic clades using published phylogenies, each nested clade was treated as a separate instance of genome duplication. In cases where this was not possible, all polyploids regardless of level were treated as a single instance of genome duplication in that clade. While this can result in multiple genome duplication events in related clades being treated as a single occurrence this is unlikely to bias our findings, as the clade is still inferred to be exclusively polyploid. Any errors are also equally likely to result in polyploid clades nested within non-polyploids as they are non-polyploids nested within polyploid clades.

Identifying sister clade pairs

Phylogenies in the published literature spanning 1975 to 2018 were used to identify the putative sister clades of our compiled list of polyploid clades. In cases where multiple phylogenies were found, the most recently published one was preferred. The only exception to this were cases where polyploid clades were part of a polytomy, in which case the most recently published phylogeny which resolved that polytomy was used to resolve that polytomy. Cases where it was impossible to resolve a polytomy containing the polyploid clade and other non-polyploid clades were discounted. In cases where the sister group of a polyploid clade was also found to be polyploid, the least inclusive clade containing both polyploidy groups was defined as the new polyploid clade and the sister clade to that new clade used. This process was repeated until there were no documented instances of polyploidy in the sister clade of the larger polyploid clade. All clade pairs therefore consisted of one polyploid clade which was inferred to contain an increase in

ploidy level and one non-polyploid clade that was inferred to not contain an increase in ploidy level. As clades were evaluated as comparable if they were phylogenetic sister clades, groups of different taxonomic ranks could be compared, although as taxonomy generally agreed with phylogeny, this was rare. More commonly, multiple taxonomic groups were contained within one or more of the clades being compared (e.g. several genera being compared to one genus) although as both clades originate from the same node in the tree (i.e. of the same phylogenetic level) we saw these comparisons as perfectly valid. This resulted in a dataset of 356 pairs of equivalent sister clades (Supplementary Table 2).

Species richness estimates

The number of species in each clade was estimated using online biodiversity databases. For vertebrates, FishBase (<http://www.fishbase.org/>)¹³⁷ was used for fish clades, AmphibiaWeb (<https://amphibiaweb.org/>)¹³⁸ for lissamphibian taxa, The Reptile Database (<http://www.reptile-database.org/>)¹³⁹ for reptile groups, Avibase (<https://avibase.bsc-eoc.org/>)¹⁴⁰ for birds, and Mammal Species of the World¹⁴¹ for mammals. For invertebrates, the Integrated Taxonomic Information System (<https://www.itis.gov/>)¹⁴², BioLib (<https://www.biolib.cz/>)¹⁴³ and Systema Dipterorum¹⁴⁴ were used for insects, the World Register of Marine Species (<http://www.marinespecies.org/>)¹⁴⁵ and the Integrated Taxonomic Information System (<https://www.itis.gov/>)¹⁴² were used for annelids, nematodes and crustaceans and the Worldwide Mollusc Species Database (<http://www.bagniliggia.it/WMSD/WMSDhome.htm>)¹⁴⁶ were used for molluscs. All plant species richness numbers were taken from The Plant List (<http://www.theplantlist.org/>)¹⁴⁷. For a minority of clades diversity estimates were also made using the source publications, which are listed for each clade. Only accepted species names were counted, known synonymies and taxa of unresolved status were not included. Extinct taxa assigned to the clades in the dataset were included in species counts using the Fossilworks portal of the Paleontology Database (<http://fossilworks.org/>)¹⁴⁸, with clade names and “fossil” as keywords. Where possible, fossil data from the Paleontology Database was checked with data in source publications to ensure species counts were as accurate as possible.

Statistical analysis

All statistical analyses were carried out using core functions in R unless otherwise stated. Paired Wilcoxon signed rank tests were carried out on the species counts of polyploid vs. non-polyploid clades in order to determine whether the number of species in polyploid clades was significantly greater than in non-polyploid clades. Separate tests were performed on the whole dataset, as well as plants and animals and major clades and grades of organism within them that were represented by at least 5 clade pairs. Wilcoxon tests were chosen as species non-normally distributed (their distributions had long tails) and were paired to account for the non-independence of sister clades in each pair. The pairwise comparison of sister groups ensures phylogenetic independence of each pair, comparing the difference between them rather than how similar they are due to phylogenetic relatedness.

In order to account for the possibility that the nesting of clades of different taxonomic ranks introduces non-independence into our data which could bias statistical tests, we performed a separate, identical, statistical

analysis using only the genera in our dataset. As the pairs of genera are all of the same taxonomic rank they cannot be nested and therefore each pair constitutes an independent test of the hypothesis. The effects of genome duplication on diversification within clades could manifest explosively as rapid speciation within a short period of time, or as slower diversification over much longer periods of time. The impacts of polyploidy on diversity could therefore be strongly or weakly scale-dependent, depending on the extent to which either of these extremes dominates the pattern. In order to determine whether taxonomic rank influenced differences in diversity between polyploid and non-polyploid clades, paired two-tailed Wilcoxon signed-rank tests were also carried out on the sample of clades of higher taxonomic rank, as well as individual taxonomic ranks represented by at least 5 clades. In addition, sign tests were implemented in R using the package BSDA to test for whether there were consistent differences in species richness between polyploid and non-polyploid groups.

Data Availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Declarations

Acknowledgements

We would like to thank Araxi Urrutia & Atehualpa Castillo-Morales for their feedback and input regarding the compilation of the data to plotting the data as well as Polly Russell and Catherine Klein for useful discussions. This research was supported by the John Templeton Foundation grants 43915 and 61408.

Author Information

Affiliations

Milner Centre for Evolution, University of Bath, Bath, UK

Jack W. Oyston & Matthew A. Wills

Vertebrates Division, Life Sciences, Natural History Museum, London, UK

Mark Wilkinson

School of Life Sciences, College of Science, University of Lincoln

Marcello Ruta

Contributions

J.W.O. and M.A.W. conceived the study and developed the methods and theory. J.W.O. compiled the data on species richness and polyploidy in clades and identified pairs of polyploid and non-polyploid sister clade

pairs from the literature. J.W.O carried out statistical analyses in R and produced the figures and tables. M.W. and M.R. helped develop the structure of the paper, contributed to discussion and helped develop analyses.

Corresponding authors

Correspondence to Jack W. Oyston or Matthew A. Wills.

Ethics declarations

Competing interests

The authors declare no competing interests.

References

1. Stork, N. E. How Many Species of Insects and Other Terrestrial Arthropods Are There on Earth? *Annual Review of Entomology* **63**, 31–45, doi:10.1146/annurev-ento-020117-043348 (2018).
2. Koenemann, S., Hoenemann, M. & Stemme, T. *World Remipedia Database*, <https://www.marinespecies.org/remipedia> (2021).
3. Schwentner, M., Combosch, D. J., Nelson, J. P. & Giribet, G. A phylogenomic solution to the origin of insects by resolving crustacean-hexapod relationships. *Current Biology* **27**, 1818–1824. e1815 (2017).
4. Lozano-Fernandez, J. *et al.* Pancrustacean evolution illuminated by taxon-rich genomic-scale data sets with an expanded remipede sampling. *Genome biology and evolution* **11**, 2055–2070 (2019).
5. von Reumont, B. M. *et al.* Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the possible sister group of Hexapoda. *Molecular biology and evolution* **29**, 1031–1045 (2012).
6. Wiens, J. J. & Dykhuizen, D. The Causes Of Species Richness Patterns Across Space, Time, And Clades And The Role Of “Ecological Limits”. *The Quarterly review of biology* **86**, 75–96, doi:10.1086/659883 (2011).
7. Stein, A., Gerstner, K. & Kreft, H. Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecology letters* **17**, 866–880 (2014).
8. Carnicer, J. & Díaz-Delgado, R. Geographic differences between functional groups in patterns of bird species richness in North America. *Acta Oecologica* **33**, 253–264 (2008).
9. Yager, J. Remipedia, a new class of Crustacea from a marine cave in the Bahamas. *Journal of Crustacean Biology*, 328–333 (1981).
10. Chothia, C., Gough, J., Vogel, C. & Teichmann, S. A. Evolution of the protein repertoire. *Science* **300**, 1701–1704 (2003).
11. Gough, J., Karplus, K., Hughey, R. & Chothia, C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *Journal of Molecular Biology* **313**, 903–919 (2001).

12. Wang, Y., Gu, X., Wang, Y. & Gu, J. Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolu. *Nature genetics* **31**, 205–209, doi:10.1038/ng902 (2002).
13. Teichmann, S. A. & Babu, M. M. Gene regulatory network growth by duplication. *Nature genetics* **36**, 492–496, doi:10.1038/ng1340 (2004).
14. Castillo-Morales, A., Monzón-Sandoval, J., Urrutia, A. O. & Gutiérrez, H. Increased brain size in mammals is associated with size variations in gene families with cell signalling, chemotaxis and immune-related functions. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **281**, 20132428–20132428, doi:10.1098/rspb.2013.2428 (2014).
15. Castillo-Morales, A. *et al.* Neocortex expansion is linked to size variations in gene families with chemotaxis, cell – cell signalling and immune response functions in mammals. *Open Biology* **6** (2016).
16. Zhang, J. Evolution by gene duplication: an update. *Trends in Ecology and Evolution* **18**, 292–298, doi:10.1016/S0169-5347(03)00033-8 (2003).
17. Bennett, M. D. Perspectives on polyploidy in plants - ancient and neo. *Biological Journal of the Linnean Society* **82**, 411–423, doi:10.1111/j.1095-8312.2004.00332.x (2004).
18. Ohno, S. *Evolution by gene duplication*. (Springer, 1970).
19. Van De Peer, Y., Maere, S. & Meyer, A. The evolutionary significance of ancient genome duplications. *Nature Reviews Genetics* **10**, 725–732, doi:10.1038/nrg2600 (2009).
20. Soltis, P. S. & Soltis, D. E. Ancient WGD events as drivers of key innovations in angiosperms. *Current Opinion in Plant Biology* **30**, 159–165, doi:10.1016/j.pbi.2016.03.015 (2016).
21. Li, Z. *et al.* Early genome duplications in conifers and other seed plants. *Science Advances* **1**, e1501084-e1501084, doi:10.1126/sciadv.1501084 (2015).
22. Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617–624, doi:10.1038/nature02424 (2004).
23. Wolfe, K. H. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–713, doi:10.1038/42711 (1997).
24. Dehal, P. & Boore, J. L. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biology* **3**, doi:10.1371/journal.pbio.0030314 (2005).
25. Donoghue, P. C. J. & Purnell, M. A. Genome duplication, extinction and vertebrate evolution. *Trends in Ecology and Evolution* **20**, 312–319, doi:10.1016/j.tree.2005.04.008 (2005).
26. Ramsey, J. & Schemske, D. W. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* **29**, 467–501, doi:10.1146/annurev.ecolsys.29.1.467 (1998).
27. Wagner, W. H. & Wagner, F. S. *Polyploidy in pteridophytes*. (Springer, 1980).
28. Adams, K. L. & Wendel, J. F. Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* **8**, 135–141, doi:10.1016/j.pbi.2005.01.001 (2005).
29. Soltis, D. E., Soltis, P. S. & Tate, J. A. Advances in the study of polyploidy since Plant Speciation. *New Phytologist* **161**, 173–191, doi:10.1046/j.1469-8137.2003.00948.x (2004).

30. Otto, S. P. & Whitton, J. Polyploid incidence and evolution. *Annual Review of Genetics* **34**, 401–437 (2000).
31. Wood, T. E. *et al.* The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 13875–13879, doi:10.1073/pnas.0811575106 (2009).
32. Levy, A. A. The impact of polyploidy on grass genome evolution. *Plant Physiology* **130**, 1587–1593, doi:10.1104/pp.015727 (2002).
33. Grivet, L. *et al.* RFLP mapping in cultivated sugarcane (*Saccharum* spp.): Genome organization in a highly polyploid and aneuploid interspecific hybrid. *Genetics* **142**, 987–1000 (1996).
34. Raboin, L. M., Pauquet, J., Butterfield, M., D'Hont, A. & Glaszmann, J. C. Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. *Theoretical and Applied Genetics* **116**, 701–714, doi:10.1007/s00122-007-0703-1 (2008).
35. Town, C. D. *et al.* Comparative Genomics of *Brassica oleracea* and *Arabidopsis thaliana* Reveal Gene Loss, Fragmentation, and Dispersal after Polyploidy. *The Plant Cell* **18**, 1348–1359, doi:10.1105/tpc.106.041665 (2006).
36. Cannon, S. B. *et al.* Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. *Molecular Biology and Evolution* **32**, 193–210, doi:10.1093/molbev/msu296 (2015).
37. Marcussen, T. *et al.* Inferring species networks from gene trees in high-polyploid North American and Hawaiian violets (*Viola*, *Violaceae*). *Systematic Biology* **61**, 107–126, doi:10.1093/sysbio/syr096 (2012).
38. Hedrén, M., Nordström, S., Persson Hovmalm, H. A., Pedersen, H. Æ. & Hansson, S. Patterns of polyploid evolution in Greek marsh orchids (*Dactylorhiza*; *Orchidaceae*) as revealed by allozymes, AFLPs, and plastid DNA data. *American Journal of Botany* **94**, 1205–1218, doi:10.3732/ajb.94.7.1205 (2007).
39. Beçak, M. L., Denaro, L. & W, B. Polyploidy and mechanisms of karyotypic diversification in Amphibia. *Cytogenetics* **9**, 225–238 (1970).
40. Haddad, C. F. B., Pombal, J. P., Jr. & Batistic, R. F. Natural hybridization between diploid and tetraploid species of leaf-frogs, genus *Phyllomedusa* (Amphibia). *Journal of Herpetology* **28**, 425–430 (1994).
41. Mable, B. K., Alexandrou, M. A. & Taylor, M. I. Genome duplication in amphibians and fish: An extended synthesis. *Journal of Zoology* **284**, 151–182, doi:10.1111/j.1469-7998.2011.00829.x (2011).
42. Braasch, I. & Postlethwait, J. H. Polyploidy in fish and the teleost genome duplication. (Springer, 2012).
43. Allendorf, F. W. *et al.* Effects of crossovers between homeologs on inheritance and population genomics in polyploid-derived salmonid fishes. *Journal of Heredity* **106**, 217–227, doi:10.1093/jhered/esv015 (2015).
44. Garcia, C., Moreira-Filho, O., Bertollo, L. A. C. & Centofante, L. B Chromosomes and Natural Triploidy in *Rhamdia* sp. (Pisces, Siluriformes, Heptapteridae). *Cytologia* **68**, 403–411, doi:10.1508/cytologia.68.403 (2003).
45. David, L., Blum, S., Feldman, M. W., Lavi, U. & Hillel, J. Recent duplication of the common carp (*Cyprinus carpio* L.) Genome as revealed by analyses of microsatellite loci. *Molecular Biology and Evolution* **20**,

- 1425–1434, doi:10.1093/molbev/msg173 (2003).
46. Lokki, Juhani, and Anssi Saura. "Polyploidy in insect evolution." *Polyploidy*. Springer, Boston, MA, 1980. 277–312.
 47. Salemaa, H. Polyploidy in the evolution of the glacial relict *Pontoporeia* spp. (Amphipoda, Crustacea). *Hereditas* **100**, 53–60, doi:10.1111/j.1601-5223.1984.tb00104.x (1984).
 48. Lee, T. Polyploidy and meiosis in the freshwater clam *Sphaerium striatinum* (Lamarck) and chromosome numbers in the Sphaeriidae (Bivalvia, Veneroidea). *Cytologia* **64**, 247–252, doi:10.1508/cytologia.64.247 (1999).
 49. Ehrenreich, I. M. Evolution after genome duplication. *Science* **368**, 1424–1425, doi:doi:10.1126/science.abc1796 (2020).
 50. Wagner, A. Selection and gene duplication: a view from the genome. *Genome biology* **3**, reviews1012-reviews1012 (2002).
 51. Soltis, P. S., Liu, X., Marchant, D. B., Visger, C. J. & Soltis, D. E. Polyploidy and novelty: Gottlieb's legacy. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130351 (2014).
 52. Mable, B. K. *et al.* Adding complexity to complexity: gene family evolution in polyploids. *Frontiers in Ecology and Evolution* **6**, 114 (2018).
 53. Purugganan, M. D., Rounsley, S. D., Schmidt, R. J. & Yanofsky, M. F. Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. *Genetics* **140**, 345–356, doi:10.1093/genetics/140.1.345 (1995).
 54. Irish, V. F. & Litt, A. Flower development and evolution: gene duplication, diversification and redeployment. *Current opinion in genetics & development* **15**, 454–460 (2005).
 55. Chanderbali, A. S., Berger, B. A., Howarth, D. G., Soltis, D. E. & Soltis, P. S. Evolution of floral diversity: genomics, genes and gamma. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20150509 (2017).
 56. Gainett, G. & Sharma, P. P. Genomic resources and toolkits for developmental study of whip spiders (Amblypygi) provide insights into arachnid genome evolution and antenniform leg patterning. *EvoDevo* **11**, 1–18 (2020).
 57. Qiao, X. *et al.* Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome biology* **20**, 1–23 (2019).
 58. Petit, C. & Thompson, J. D. Species diversity and ecological range in relation to ploidy level in the flora of the Pyrenees. *Evolutionary Ecology* **13**, 45–66, doi:10.1023/A:1006534130327 (1999).
 59. Vamosi, J. C. & Dickinson, T. a. Polyploidy and diversification: a phylogenetic investigation in Rosaceae. *International Journal of Plant Sciences* **167**, 349–358, doi:10.1086/499251 (2006).
 60. Rieseberg, L. H. & Willis, J. H. Plant speciation. *Science* **17**, 910–914, doi:10.1126/science.177.4049.600 (2007).
 61. Soltis, D. E., Buggs, R. J. a., Doyle, J. J. & Soltis, P. S. What we still don't know about polyploidy. *Taxon* **59**, 1387–1403, doi:10.2307/20774036 (2010).

62. Arrigo, N. & Barker, M. S. Rarely successful polyploids and their legacy in plant genomes. *Current Opinion in Plant Biology* **15**, 140–146, doi:10.1016/j.pbi.2012.03.010 (2012).
63. Mayrose, I. *et al.* Recently formed polyploid plants diversify at lower rates. *Science* **333**, 1257–1257, doi:10.1126/science.1207205 (2011).
64. Lutz, A. M. A preliminary note on the chromosomes of *Oenothera Lamarckiana* and one of its mutants, *O. gigas*. *Science* **26**, 151–152 (1907).
65. Stebbins, G. L. *Variation and evolution in plants*. (Columbia University Press, 1950).
66. Winge, O. The chromosomes. Their numbers and general importance. *Comptes Rendus des Travaux du Laboratoire Carlsberg* **13**, 131–175 (1917).
67. Stebbins, G. L. *Chromosomal evolution in higher plants*. (Edward Arnold, 1971).
68. Wagner, W. H. Biosystematics and evolutionary noise. *Taxon* **19**, 146–151 (1970).
69. Barker, M. S., Arrigo, N., Baniaga, A. E., Li, Z. & Levin, D. A. On the relative abundance of autopolyploids and allopolyploids. *New Phytologist* **210**, 391–398, doi:10.1111/nph.13698 (2016).
70. Soltis, D. E., Visger, C. J., Marchant, D. B. & Soltis, P. S. Polyploidy: pitfalls and paths to a paradigm. *American Journal of Botany* **103**, 1146–1166, doi:10.3732/ajb.1500501 (2016).
71. Crepet, W. L. & Niklas, K. J. Darwin's second 'abominable mystery': why are there so many angiosperm species? *American journal of botany* **96**, 366–381, doi:10.3732/ajb.0800126 (2009).
72. Leitch, I. J. & Bennett, M. D. Polyploidy in angiosperms. *Trends in Plant Science* **5** (1997).
73. Stull, G. W. *et al.* Gene duplications and genomic conflict underlie major pulses of phenotypic evolution in gymnosperms. *bioRxiv* (2021).
74. Ahuja, M. R. Polyploidy in gymnosperms: revisited. *Silvae Genetica* **54**, 59–69, doi:10.1515/sg-2005-0010 (2005).
75. Rothleitner, J. J., Friddle, M. W. & Contreras, R. N. Ploidy levels, relative genome sizes, and base pair composition in Magnolia. *Journal of the American Society for Horticultural Science* **135**, 533–547, doi:10.21273/JASHS03776-16 (2010).
76. Dauphin, B., Grant, J. R., Farrar, D. R. & Rothfels, C. J. Rapid allopolyploid radiation of moonwort ferns (*Botrychium*; *Ophioglossaceae*) revealed by PacBio sequencing of homologous and homeologous nuclear regions. *Molecular Phylogenetics and Evolution* **120**, 342–353, doi:10.1016/j.ympev.2017.11.025 (2018).
77. Schneider, H. *et al.* Neo- and Paleopolyploidy contribute to the species diversity of *Asplenium* – the most species-rich genus of ferns. *Journal of Systematics and Evolution* **55**, 353–364, doi:10.1111/jse.12271 (2017).
78. Taylor, J. S., Braasch, I., Frickey, T. & Meyer, A. Genome duplication, a trait shared by 22 000 species of ray-finned fish. *Genome Research* **13**, 382–390, doi:10.1101/gr.640303.1 (2003).
79. Burke, A. C., Nelson, C. E., Morgan, B. A. & Tabin, C. J. Hox genes and the evolution of vertebrate axial morphology. *Development* **121**, 333–346, doi:7768176 (1995).
80. Kassahn, K. S., Dang, V. T., Wilkins, S. J., Perkins, A. C. & Ragan, M. A. Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. *Genome*

- Research **19**, 1404–1418, doi:10.1101/gr.086827.108.1404 (2009).
81. Braasch, I., Brunet, F., Volff, J. N. & Schartl, M. Pigmentation pathway evolution after Whole-Genome Duplication in fish. *Genome Biology and Evolution* **1**, 479–493, doi:10.1093/gbe/evp050 (2009).
 82. Renz, A. J., Meyer, A. & Kuraku, S. Revealing less derived nature of cartilaginous fish genomes with their evolutionary time scale inferred with nuclear genes. *PLoS ONE* **8**, doi:10.1371/journal.pone.0066400 (2013).
 83. Ravi, V. & Venkatesh, B. Rapidly evolving fish genomes and teleost diversity. *Current Opinion in Genetics and Development* **18**, 544–550, doi:10.1016/j.gde.2008.11.001 (2008).
 84. Evans, B. J., Kelley, D. B., Tinsley, R. C., Melnick, D. J. & Cannatella, D. C. A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. *Molecular Phylogenetics and Evolution* **33**, 197–213, doi:10.1016/j.ympev.2004.04.018 (2004).
 85. Martino, A. L. & Sinsch, U. Speciation by polyploidy in *Odontophrynus americanus*. *Journal of Zoology* **257**, 67–81, doi:10.1017/S0952836902000663 (2002).
 86. Ptacek, M. B., Gerhardt, H. C. & Sage, R. D. Speciation by polyploidy in treefrogs: multiple origins of the tetraploid, *Hyla versicolor*. *Evolution* **48**, 898–908 (1994).
 87. Clinton, M. Sex determination and gonadal development: a bird's eye view. *The Journal of Experimental Zoology* **281**, 457–465 (1998).
 88. Tiersch, T. R., Beck, M. L. & Douglass, M. ZZW autotriploidy in a Blue-and-Yellow Macaw. *Genetica* **84**, 209–212, doi:10.1007/BF00127249 (1991).
 89. Suárez-Villota, E. Y. *et al.* Distribution of repetitive DNAs and the hybrid origin of the red vizcacha rat (Octodontidae). *Genome* **55**, 105–117, doi:10.1139/g11-084 (2012).
 90. Evans, B. J., Upham, N. S., Golding, G. B., Ojeda, R. A. & Ojeda, A. A. Evolution of the largest mammalian genome. *Genome Biology and Evolution* **9**, 1711–1724, doi:10.1093/gbe/evx113 (2017).
 91. Song, C. *et al.* Polyploid organisms. *Science China Life Sciences* **55**, 301–311, doi:10.1007/s11427-012-4310-2 (2012).
 92. Hughes, C. L. & Kaufman, T. C. Hox genes and the evolution of the arthropod body plan. *Evolution & Development* **4**, 459–499, doi:DOI 10.1046/j.1525-142X.2002.02034.x (2002).
 93. Lemons, D. & McGinnis, W. Genomic evolution of hox gene clusters. *Science* **313**, 1918–1922, doi:10.1126/science.1132040 (2006).
 94. Galant, R. & Carroll, S. B. Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* **415**, 910–913, doi:10.1038/nature717 (2002).
 95. Ronshaugen, M., McGinnis, N. & McGinnis, W. Hox protein mutation and macroevolution of the insect body plan. *Nature* **415**, 914–917, doi:10.1038/nature716 (2002).
 96. Gregory, T. R. & Hebert, P. D. N. Genome size estimates for some oligochaete annelids. *Canadian Journal of Zoology* **80**, 1485–1489, doi:10.1139/z02-145 (2002).
 97. Marotta, R., Crottini, A., Raimondi, E., Fondello, C. & Ferraguti, M. Alike but different: the evolution of the *Tubifex tubifex* species complex (Annelida, Clitellata) through polyploidization. *BMC Evolutionary Biology* 2014, 14-73, doi:10.1186/1471-2148-14-73 (2014).

98. Parker, E. D. & Selander, R. K. The organization of genetic diversity in the parthenogenetic lizard *Cnemidophorus tesselatus*. *Genetics* **84**, 791–805 (1976).
99. Case, T. J. Patterns of coexistence in sexual and asexual species of *Cnemidophorus* lizards. *Oecologia* **83**, 220–227, doi:10.1007/BF00317756 (1990).
100. Cullum, A. J. Phenotypic variability of physiological traits in populations of sexual and asexual whiptail lizards (genus *Cnemidophorus*). *Evolutionary Ecology Research* **2**, 841–855 (2000).
101. Blanc, G., Hokamp, K. & Wolfe, K. H. A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Research* **13**, 137–144, doi:10.1101/gr.751803 (2003).
102. Gregory, T. R. & Mable, B. K. Polyploidy in animals. *The evolution of the genome*. (ed T. Ryan Gregory) Vol. 171 427-517 (Elsevier, San Diego, 2005).
103. Tate, J. A., Soltis, D. E. & Soltis, P. S. Polyploidy in plants. *The evolution of the genome*. (ed T. Ryan Gregory) Vol. 171 371-426 (Elsevier Academic Press, 2005).
104. Comai, L. The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* **6**, 836–846, doi:10.1038/nrg1711 (2005).
105. Neiman, M., Beaton, M. J., Hessen, D. O., Jeyasingh, P. D. & Weider, L. J. Endopolyploidy as a potential driver of animal ecology and evolution. *Biological Reviews* **92**, 234–247, doi:10.1111/brv.12226 (2017).
106. Jaekel, M. & Wake, D. B. Developmental processes underlying the evolution of a derived foot morphology in salamanders. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 20437–20437, doi:10.1073/pnas.0710216105 (2007).
107. Flemming, A. J., Shen, Z.-Z., Cunha, A., Emmons, S. W. & Leroi, A. M. Somatic polyploidization and cellular proliferation drive body size evolution in nematodes. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5285–5290, doi:10.1073/pnas.97.10.5285 (2000).
108. Henery, C. C., Bard, J. B. L. & Kaufman, M. H. Tetraploidy in mice, embryonic cell number, and the grain of the developmental map. *Developmental Biology* **152**, 233–241, doi:10.1016/0012-1606(92)90131-Y (1992).
109. Akanno, E. C. *et al.* Modelling heterotic effects in beef cattle using genome-wide SNP-marker genotypes. doi:10.1093/jas/skx002/4823823 (2018).
110. Guo, M., Davis, D. & Birchler, J. A. Dosage effects on gene expression in a maize ploidy series. *Genetics* **142**, 1349–1355, doi:10.1016/0168-9525(96)81463-6 (1996).
111. Standish, K., Allen, J. R. & Stanley, J. G. Reproductive sterility in polyploid Brook Trout, *Salvelinus fontinalis*. *Transactions of the American Fisheries Society* **107**, 473–478 (1978).
112. Matzke, M. A., Mette, M. F., Kanno, T. & Matzke, A. J. M. Does the intrinsic instability of aneuploid genomes have a causal role in cancer? *Trends in Genetics* **19**, 253–256, doi:10.1016/S0168-9525(03)00057-X (2003).
113. Fechheimer, N. S. Origins of heteroploidy in chicken embryos. *Poultry Science* **60**, 1365–1371 (1981).
114. Leitch, A. R. & Leitch, I. J. Genomic plasticity and the diversity of polyploid plants. *Science* **320**, 481–483, doi:10.1126/science.1153585 (2008).

115. Miller, J. S. & Venable, D. L. Polyploidy and the evolution of gender dimorphism in plants. *Science* **289**, 2335–2338, doi:10.1126/science.289.5488.2335 (2000).
116. Bierzychudek, P. Patterns in plant parthenogenesis. *Experientia* **41**, 1255–1264 (1985).
117. Soltis, P. S. & Soltis, D. E. The role of hybridization in plant speciation. *Annual Review of Plant Biology* **60**, 561–588, doi:10.1146/annurev.arplant.043008.092039 (2009).
118. Charlesworth, D. Evolution: An exception that proves the rule. *Current Biology* **11**, R13-R15, doi:https://doi.org/10.1016/S0960-9822(00)00035-X (2001).
119. Haeckel, E. H. *Generelle Morphologie der Organismen allgemeine Grundzuge der organischen Formen-Wissenschaft, mechanisch begrundet durch die von Charles Darwin reformirte Descendenz-Theorie von Ernst Haeckel: Allgemeine Entwicklungsgeschichte der Organismen kritische G.* (Verlag von Georg Reimer, 1866).
120. Olson, E. N. Gene regulatory networks in the evolution and development of the heart. *Science* **313**, 1922–1927 (2006).
121. Taft, R. J., Pheasant, M. & Mattick, J. S. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **29**, 288–299 (2007).
122. Kohsokabe, T. & Kaneko, K. Dynamical systems approach to evolution–development congruence: Revisiting Haeckel's recapitulation theory. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **338**, 62–75 (2022).
123. Haeckel, E. H. *Anthropogenie: Keimes-und Stammes-Geschichte des Menschen.* (Engelmann, 1874).
124. Olsson, L., Levit, G. S. & Hoßfeld, U. The “Biogenetic Law” in zoology: from Ernst Haeckel's formulation to current approaches. *Theory in Biosciences* **136**, 19–29, doi:10.1007/s12064-017-0243-4 (2017).
125. Carroll, S. B. Evo-Devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36, doi:10.1016/j.cell.2008.06.030 (2008).
126. Williams, G. C. Vol. 11 398–398 (1957).
127. Hornstein, E. & Shomron, N. Canalization of development by micrnas. *Nature Genetics* **38**, S20-S20, doi:10.1038/ng1803 (2006).
128. Peterson, K. J., Dietrich, M. R. & McPeck, M. A. MicroRNAs and metazoan macroevolution: Insights into canalization, complexity, and the Cambrian explosion. *BioEssays* **31**, 736–747, doi:10.1002/bies.200900033 (2009).
129. Reznick, D. N., Rodd, H. F. & Cardenas, M. Life-history evolution in guppies (*Poecilia reticulata*: Poeciliidae). IV. Parallelism in life-history phenotypes. *The American Naturalist* **147**, 319–338 (1996).
130. Sucena, E., Delon, I., Jones, I., Payre, F. & Stern, D. L. Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature* **424**, 935–938, doi:10.1038/nature01768 (2003).
131. Yoon, H.-S. & Baum, D. a. Transgenic study of parallelism in plant morphological evolution. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 6524–6529, doi:10.1073/pnas.0401824101 (2004).
132. Schluter, D. Adaptive radiation along genetic lines of least resistance. *Evolution* **5**, 1766–1774 (1996).

133. Marchant, A. *et al.* The role of structural pleiotropy and regulatory evolution in the retention of heteromers of paralogs. *Elife* **8**, e46754 (2019).
134. Crow, K. D. & Wagner, G. P. What is the role of genome duplication in the evolution of complexity and diversity? *Molecular biology and evolution* **23**, 887–892 (2005).
135. Martin, S. L. & Husband, B. C. Whole genome duplication affects evolvability of flowering time in an autotetraploid plant. (2012).
136. Payne, J. L. & Wagner, A. The causes of evolvability and their evolution. *Nature Reviews Genetics* **20**, 24–38 (2019).
137. Froese, R. & Pauly, D. FishBase, <http://www.fishbase.org/> (2019).
138. AmphibiaWeb, <<https://amphibiaweb.org/>> (2018).
139. Uetz, P., Freed, P. & Hošek, J. The Reptile Database, <http://www.reptile-database.org> (2019).
140. Lepage, D., Vaidya, G. & Guralnick, R. Avibase - A database system for managing and organizing taxonomic concepts. *ZooKeys* **135**, 117–135, doi:10.3897/zookeys.420.7089 (2014).
141. Wilson, D. E. & Reeder, D. M. *Mammal Species of the World. A Taxonomic and Geographic Reference* (3rd ed). 3rd edn, (Johns Hopkins University Press, 2005).
142. The Integrated Taxonomic Information System. <http://www.itis.gov/> (2018).
143. Zicha, O. BioLib, <https://www.biolib.cz/en/> (2018).
144. Pape, T. & Thompson, F. C. *Systema Dipteroorum*, Version 1.5, <http://www.diptera.org/> (2013).
145. WoRMS Editorial Board. *World Register of Marine Species*, <<http://www.marinespecies.org>> (2018).
146. Galli, C. WMSDB - Worldwide Mollusc Species Data Base, <http://www.bagniliggia.it/WMSD/WMSDhome.htm> (2016).
147. The Plant List Version 1.1, <http://www.theplantlist.org> (2018).
148. Alroy, J. Fossilworks, <http://fossilworks.org> (2013).
149. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>.

Tables

Table 1: Differences in the species richness of polyploid and non-polyploid clades.

Dataset	Number of clade pairs	Average no. of species (polyploid clades)		Average no. of species (non-polyploid clades)		Wilcoxon signed-rank V	P-value
		Mean	Median	Mean	Median		
All	356	406	52	93	15	45769	6.734x10 ⁻¹⁶
Animals	153	649	39	67	15	8825	9.807ex10 ⁻⁰⁹
Plants	203	220	70	112	15	14468	1.072x10 ⁻⁰⁸
Vertebrates	91	83	25	44	12	3025	2.895x10 ⁻⁰⁵
Invertebrates	62	1489	60	102	25	1542	7.445x10 ⁻⁰⁵
Angiosperms	128	257	57	133	18	5501.5	5.018x10 ⁻⁰⁵
Non-angiosperms	75	160	75	76	11	2155	3.595x10 ⁻⁰⁵
Tetrapods	49	36	14	34	10	767.5	0.06631
Fish	42	169	77	82	14	720	0.0001802
Lissamphibians	29	32	15	34	9	285.5	0.0618
Reptiles	17	48	24	38	15	91	0.5074
Insects	32	2760	55	134	21	386.5	0.0225
Annelids	13	174	48	97	32	67	0.1421
Crustaceans	6	59	26	19	3	21	0.03552
Molluscs	10	135	105	63	60	52.5	0.01246
Magnolids + Chloranthales	9	1365	272	388	10	38	0.07422
Dicots	91	187	52	130	24	2654.5	0.003779
Monocots	27	128	36	61	15	251.5	0.05512
Gymnosperms	6	53	49	12	6	15	0.05906
Ferns	69	169	77	82	14	1852	0.0001179

Number of clade pairs and average number of species within each clade, as well as two-tailed paired Wilcoxon signed-rank tests on the clade pairs of all taxonomic ranks, for the entire dataset (All) as well as clade pairs in each subgroup.

Table 2: Differences in the species richness of polyploid and non-polyploid clades.

Dataset	Number of clade pairs	Average no. of species (polyploid clades)		Average no. of species (non-polyploid clades)		Wilcoxon signed-rank V	P-value
		Mean	Median	Mean	Median		
All	356	406	52	93	15	45769	6.734x10 ⁻¹⁶
Above Genus	35	2952	251	288	54	493	0.002833
Families	28	3459	251	322	57	311	0.0127
Genera	321	129	46	70	13	36829	5.645x10 ⁻¹⁴

Two-tailed paired Wilcoxon signed-rank tests performed on polyploid and non-polyploid data subsets of different taxonomic ranks. All: the entire dataset of clades at all taxonomic ranks, Above Genus: all clades of a taxonomic rank above higher than genus.

Table 3: Differences in the species richness of polyploid and non-polyploid genera.

Dataset	Number of clade pairs	Average no. of species (polyploid genera)		Average no. of species (non-polyploid genera)		Wilcoxon signed-rank V	P-value
		Mean	Median	Mean	Median		
All	321	129	46	70	13	36829	5.645x10 ⁻¹⁴
Animals	129	87	30	37	12	6150	5.209x10 ⁻⁰⁷
Plants	192	157	66	92	15	12928	2.351x10 ⁻⁰⁸
Vertebrates	74	46	16	24	10	1941.5	0.0004334
Invertebrates	55	142	46	55	16	1200.5	0.0003143
Angiosperms	117	156	52	102	17	4556.5	0.0001302
Non-angiosperms	75	160	75	76	11	2155	3.595x10 ⁻⁰⁵
Tetrapods	48	36	14	34	10	767.5	0.03166
Fish	26	64	43	24	11	263	0.00712
Lissamphibians	28	33	14	17	8	285.5	0.02105
Reptiles	17	48	24	38	15	91	0.5074
Insects	30	195	46	78	17	338.5	0.02996
Annelids	10	83	31	29	22	40	0.221
Crustaceans	6	59	26	19	3	21	0.03552
Molluscs	8	100	80	37	38	34	0.02344
Magnolids + Chloranthales	3	260	272	12	10	6	0.25
Dicots	87	161	52	118	23	2457.5	0.002726
Monocots	27	128	36	61	15	251.5	0.05512
Gymnosperms	6	53	49	12	6	15	0.05906

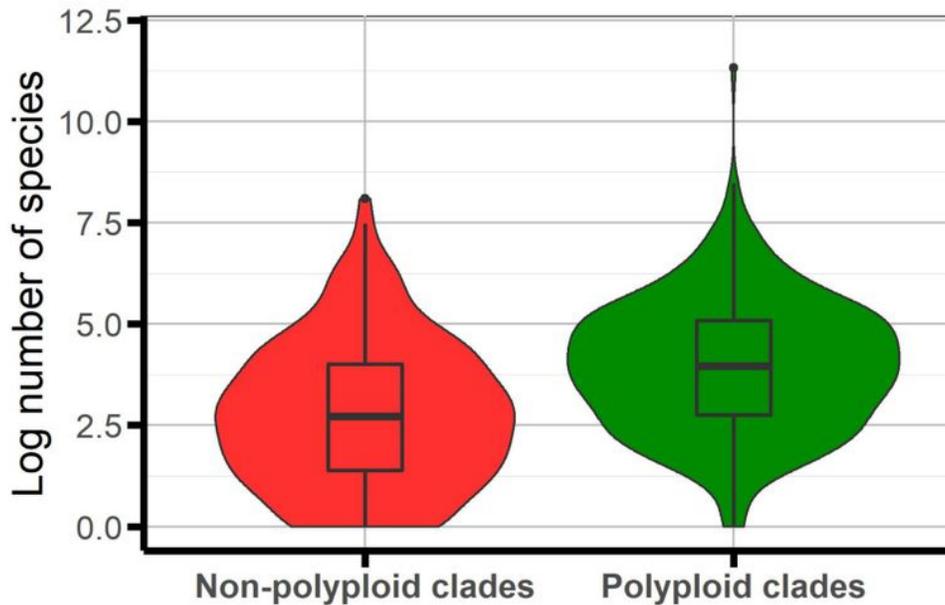
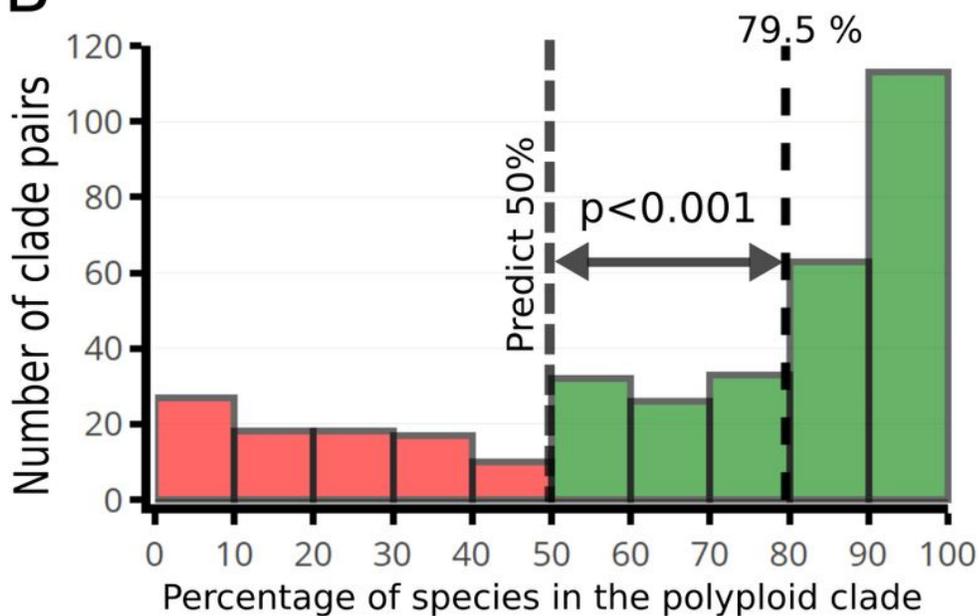
Two-tailed paired Wilcoxon signed-rank tests on pairs of genera, for the entire dataset (All) as well as clade pairs in each subgroup.

Table 4: Sign tests of differences in species richness of polyploid and non-polyploid clades

Dataset	All Levels			Genera Only		
	No. of clade pairs (total with difference)	No. of clade pairs (polyploids higher)	p-value	No. of clade pairs (total with difference)	No. of clade pairs (polyploids higher)	p-value
All	349	260	$< 2.2 \times 10^{-16}$	314	233	2.2×10^{-16}
Animals	151	111	6.474×10^{-9}	127	92	4.361×10^{-7}
Plants	198	149	6.284×10^{-13}	187	141	2.057×10^{-12}
Vertebrates	89	66	5.69×10^{-6}	72	52	0.0002077
Invertebrates	62	45	0.0004971	55	40	0.001016
Angiosperms	124	91	1.859×10^{-7}	113	83	6.307×10^{-7}
Non-angiosperms	74	58	9.675×10^{-7}	74	58	9.675×10^{-7}
Tetrapods	49	32	0.0293	49	32	0.0293
Fish	41	34	2.532×10^{-5}	25	20	0.004077
Lissamphibians	28	20	0.0357	27	20	0.01916
Reptiles	17	10	0.6291	17	10	0.6291
Insects	32	20	0.2153	30	19	0.2005
Annelids	13	10	0.09229	10	8	0.1094
Crustaceans	6	6	0.03125	6	6	0.03125
Molluscs	10	9	0.02148	8	7	0.07031
Magnolids + Chloranthales	9	8	0.03906	3	3	0.25
Dicots	88	63	6.27×10^{-5}	84	61	4.077×10^{-5}
Monocots	26	19	0.02896	26	19	0.02896
Gymnosperms	5	5	0.0625	5	5	0.0625

Sign tests on differences in species richness of polyploid and non-polyploid clades, showing the numbers of clade pairs with more speciose polyploid clades compared to all clade pairs showing a difference in species richness, for the whole dataset (All levels) and for pairs of genera only (Genera only).

Figures

A**B****Figure 1****Species richness in polyloid and non-polyloid sister clades.**

A: Violin plot showing differences in log species richness between sister clades inferred to have a basal ploidy event (polyloid) and those without (non-polyloid). Boxes delimit the upper and lower quartiles of the data, while central bars are median values. **B:** Histogram of clade pairs grouped by the total percentage of species in both clades found in the polyloid clade. The red area of the plot indicates more species in the

non-polyploid clade, while the green area indicates more species in the polyploid clade. Observed median of 79.5% of species in the polyploid clade (right-hand dashed line) differs significantly ($p < 0.001$) from the null prediction that 50% of species are in the polyploid clade (left-hand dashed line).

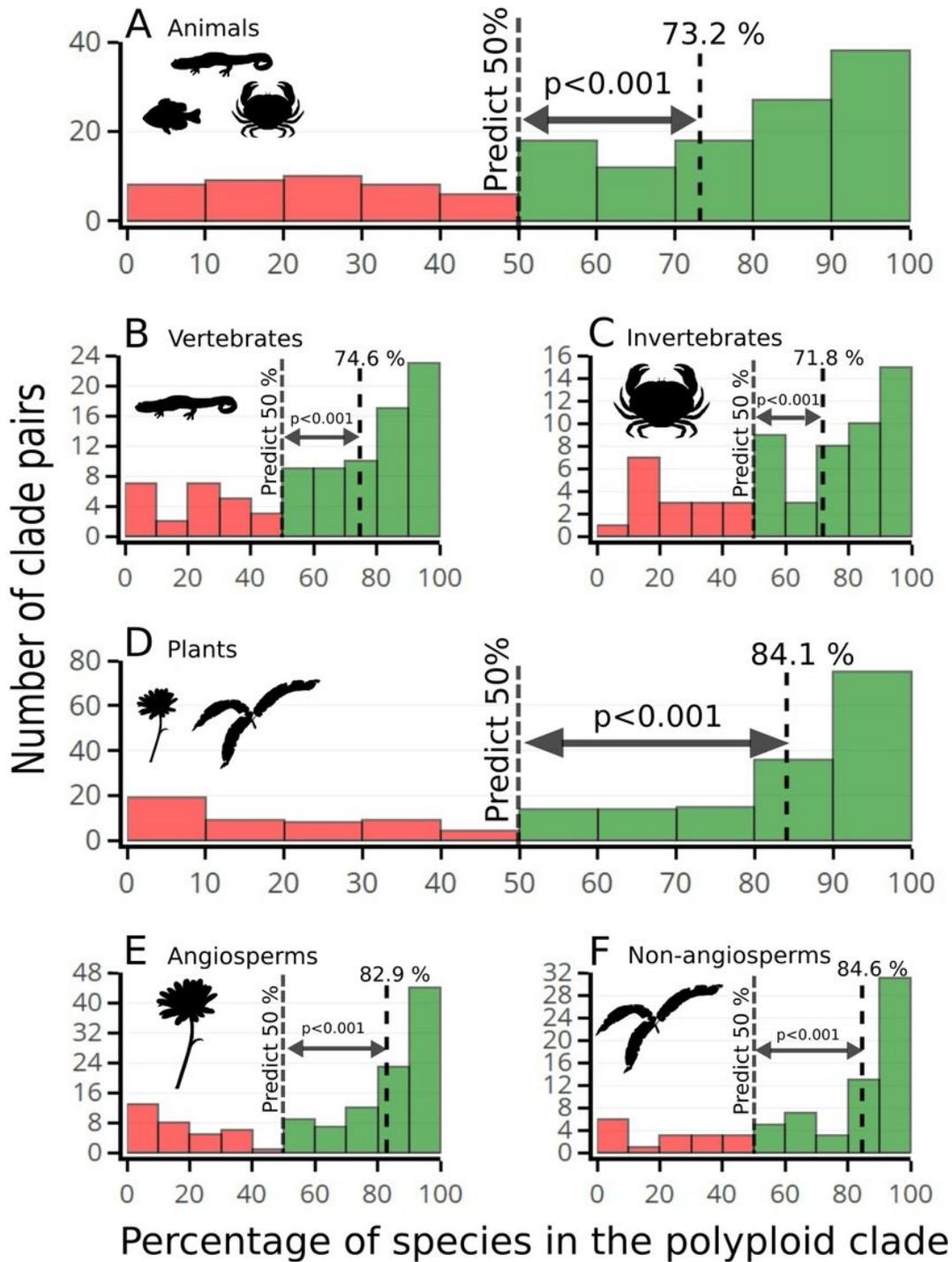


Figure 2

Relative diversity of polyploid and non-polyploid sister clades in animal and plant groups.

Histogram of clade pairs in **A**: all animals, **B**: vertebrates, **C**: invertebrates, **D**: all plants, **E**: angiosperms and **F**: non-angiosperms grouped by the total percentage of species in both clades found in the polyploid clade. The red area of the plot indicates more species in the non-polyploid clade, while the green area indicates more species in the polyploid clade. Observed medians of species numbers in the polyploid clade (right-hand dashed line) differ significantly ($p < 0.001$) from the null prediction that 50% of species are in the polyploid clade (left-hand dashed line) in all cases.

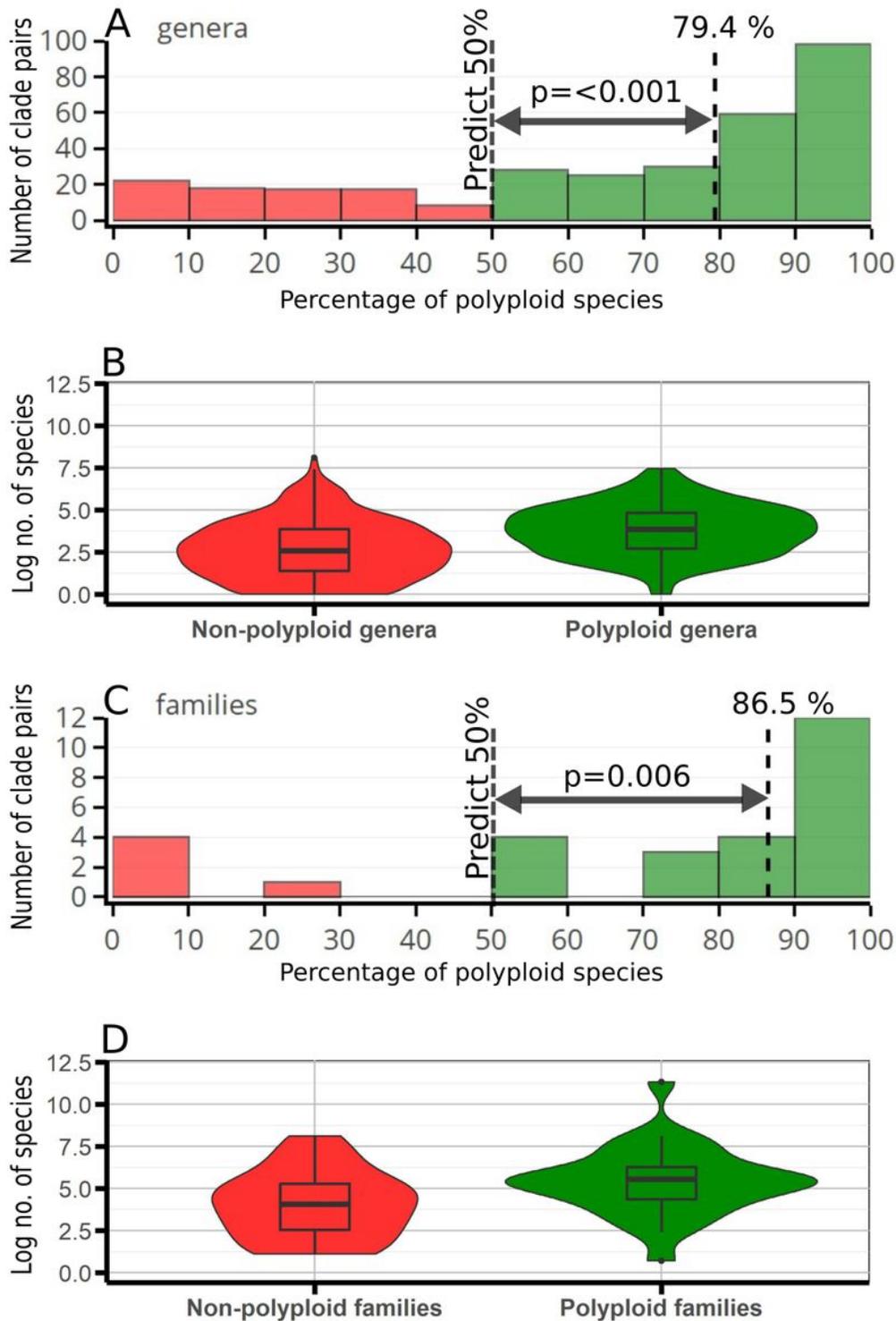


Figure 3

Species richness of polyploid and non-polyploid sister clades at different taxonomic levels.

Differences in species richness between sister genera (**A, B**) and sister families (**C, D**) inferred to have a basal ploidy event (polyploid) and those without (non-polyploid). Violin plots showing differences in log species richness between sister clades for **A**: genera and **C**: families. Histogram clade pairs of **B**: genera and **D**: families grouped by the total percentage of species found in the polyploid clade. Ranges in the red area of the plot indicate more species in the non-polyploid clade, while those in the green area indicate more species in the polyploid clade.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryDataset1.csv](#)
- [Polyploidcladestend tobemorespeciosethantheirputativesistercladesESMNat.Comms.pdf](#)
- [rs.pdf](#)