Author for correspondence:
John J. Wiens
e-mail: wiensj@email.arizona.edu

THE ROYAL SOCIETY
PUBLISHING

Multicellularity and sex helped shape the Tree of Life

Lian Chen[1,2] and John J. Wiens[2]

[1]College of Biology and the Environment, Nanjing Forestry University, Nanjing 210037, People's Republic of China
[2]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721-0088, USA

JJW, 0000-0003-4243-1127

Across the Tree of Life, there are dramatic differences in species numbers among groups. However, the factors that explain the differences among the deepest branches have remained unknown. We tested whether multicellularity and sexual reproduction might explain these patterns, since the most species-rich groups share these traits. We found that groups with multicellularity and sexual reproduction have accelerated rates of species proliferation (diversification), and that multicellularity has a stronger effect than sexual reproduction. Patterns of species richness among clades are then strongly related to these differences in diversification rates. Taken together, these results help explain patterns of biodiversity among groups of organisms at the very broadest scales. They may also help explain the mysterious preponderance of sexual reproduction among species (the 'paradox of sex') by showing that organisms with sexual reproduction proliferate more rapidly.

1. Introduction

Explaining the dramatic differences in species diversity among groups of organisms is a fundamental challenge in evolutionary biology [1]. Yet, few studies, if any, have attempted to explain richness patterns among the deepest branches of the Tree of Life. For example, animals, land plants and fungi each have approximately 1.5 million, 350 000 and 140 000 described species (figure 1), respectively. By contrast, most other major groups (e.g. bacteria, archaeans, various protist clades) have far fewer described species (figure 1), in the tens of thousands or less. What explains these striking differences in diversity?

To our knowledge, no previous studies have attempted to explain patterns of species diversity at this scale. Nevertheless, an earlier study [2] did examine patterns of species richness among eight kingdom-level clades (e.g. animals, fungi, plants, bacteria, archaeans, various protists). That study concluded that most variation in species richness among these clades (approx. 55%) was explained by differences in their rates of diversification and not their ages. Diversification rates describe how quickly richness accumulated within clades, or the rate of speciation minus the rate of extinction [3,4]. Thus, clades that are relatively young and have high extant species richness (like animals, fungi and land plants) have higher net diversification rates than those that are older or have fewer living species (which then have lower diversification rates). This pattern raises the obvious question: what trait or traits might explain these accelerated diversification rates, especially among animals, plants and fungi?

Here, we test if two traits may help explain these patterns: multicellularity and sexual reproduction. An obvious trait shared by animals, land plants and many fungi is that they are multicellular, with adult bodies consisting of many connected cells. Multicellularity has evolved many times, and is considered a major evolutionary transition [5–7]. Multicellularity is potentially important as a driver of diversification because multicellularity may be a necessary precursor to overall morphological complexity (e.g. allowing for different cell and tissue types with
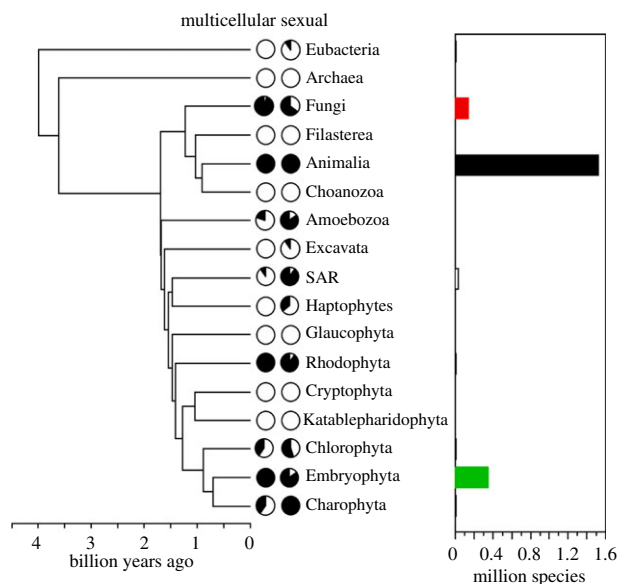
Figure 1. Summary of the 17 kingdom-level clades analysed in this study. The time-calibrated phylogeny is shown on the left. For each clade, we show the estimated proportion of species that are multicellular and the proportion with sexual reproduction (in black). The species richness of each clade (based on described species) is summarized in the graph on the right. Most clades have so few species (relative to animals, fungi and land plants) that they are not visible in this graph, but we used raw (not log-transformed) values to better illustrate the actual disparity in richness. Embryophyta corresponds to land plants. The tree is given in the electronic supplementary material, datafile S1 and the data in the electronic supplementary material, datafile S5. Note that we also tested the impacts of treating many prokaryotic clades as separate units, and of including eukaryotes only (see Results). (Online version in colour.)

different functions [6,7]). However, there have been few tests of whether diversification rates and multicellularity are correlated, and these have been at relatively small phylogenetic scales. For example, a study within cyanobacteria found that multicellularity might increase diversification rates within this group [8]. Although this result is intriguing, studies within particular clades cannot directly address whether multicellularity explains variation in diversification rates across the entire Tree of Life.

The presence of sexual reproduction might also shed light on these large-scale diversity patterns, as it is also present in the three largest kingdom-level clades (animals, fungi, land plants). Yet, the larger number of species that reproduce sexually (at least part of the time) relative to those that reproduce only asexually is a long-standing puzzle in evolutionary biology [9–13]. This puzzle is often framed in terms of the costs of sexual reproduction relative to asexual reproduction, and why sexual reproduction is maintained in sexual species [12,13]. Differences in diversification rates associated with each mode might also strongly influence the relative richness of species with each reproductive mode. To our knowledge, no studies have tested whether sexual and asexual reproduction are associated with different diversification rates across the Tree of Life. However, as for multicellularity, there have been important studies within smaller clades, such as rotifers [14]. Much of this literature has focused on the persistence of secondarily asexual lineages within ancestrally sexual clades [15,16].

Possible links between sexual reproduction and multi-cellularity have also not been tested at this deep scale.

Multicellularity has been hypothesized to precede and underlie the evolution of differentiated sexes [11,17,18], if not sexual reproduction itself.

In this study, we analyse the relationships between diversification rates, multicellularity and sexual reproduction among major clades across the Tree of Life. First, we estimate the proportions of multicellular species and sexually reproducing species in each of 17 kingdom-level clades (figure 1), based on a literature survey spanning more than 1146 papers. These 17 clades include animals, land plants, fungi, bacteria, archaeans and 12 major groups of protists and algae (we also explore subdividing some of these clades). Our definitions of these traits are given in the Material and methods, and the electronic supplementary material, appendix S1. Second, we estimate rates of diversification for each of these clades, using a standard method for higher-level taxa [19]. Third, we test for relationships between diversification rates and sexual reproduction and multicellularity using phylogenetic regression. Using this overall approach, we can address how much variance in diversification rates each variable statistically explains (both alone and in combination), not simply whether each trait has a significant effect on diversification. We also address how much variation in species richness among clades is explained by this variation in diversification rates. Fourth, we test whether sexual reproduction and multicellularity are significantly related to each other at this scale. We perform these analyses using both numbers of described species, and using projections (estimates) of species richness that suggest there may be hundreds of million (or even billions) of undescribed species, especially of bacteria [20]. These projections are described in the electronic supplementary material, appendix S2.

2. Material and methods

(a) Phylogeny

We used a previously assembled phylogeny that spanned the Tree of Life [2], with the eukaryotic portion from Parfrey et al. [21]. We initially used 17 major clades from this tree [2]. These clades were either ranked as kingdoms or were distinct clades outside the commonly recognized kingdoms. Scholl & Wiens [2] used eight major non-overlapping clades in their kingdom-level analyses: Archaea, Eubacteria, Animalia, Plantae (Embryophyta), Fungi, Amoebozoa, Excavata and the SAR clade. However, their tree also included nine additional clades that did not overlap with each other or the other eight clades. These nine clades are traditionally classified as protists and/or algae, and consisted of: Charophyta, Chlorophyta, Choanoflagellatea, Cryptophyta, Filasterea, Glaucophyta, Haptophyta, Katablepharidophyta and Rhodophyta. We pruned the tree [2] to include only one species from each clade (the choice has no impact). This tree was then used in the phylogenetic regression analyses, and is given in the electronic supplementary material, datafile S1.

Hypothetically, the Timetree of Life [22] could have been used to construct an alternative tree. However, this tree did not resolve relationships (or ages) among some relevant clades (e.g. Glaucophyta, Excavata, Rhodophyta). Furthermore, many dates were from Parfrey et al. [21], such that this source is not necessarily an alternative estimate.

Charophyta was treated as a clade here, containing Charophyceae, Chlorokybophyceae, Coleochaetophyceae and Zygnemophyceae. However, this taxon may be paraphyletic with respect to land plants [23,24]. Although this is problematic, most species in Charophyta belong to Zygnemophyceae (4107 of 4867; electronic supplementary material, datafiles S2–S4), which

Table 1. Relationships among diversification, sexuality, multicellularity and species richness. (Analyses are performed using the number of described species in each of the 17 kingdom-level clades (electronic supplementary material, datafile S5), a relatively low projected number of species for several major clades (electronic supplementary material, datafile S8) and a relatively high projected number of species (electronic supplementary material, datafile S9). Diversification rates were estimated using $\varepsilon = 0.5$. Results using alternative values are very similar (electronic supplementary material, tables S1 and S2). Lambda is the estimated phylogenetic signal for the relationship. Values in parentheses for the multiple regression model are the standardized partial regression coefficients for each independent variable. Akaike information criterion (AIC) values were often negative when the dependent variable was diversification rates. We did not compare positive and negative AIC values. Italicized p-values are significant after a table-wide, sequential Bonferroni correction.)

dependent variable	independent variable	AIC	λ	r^2	p-value
described richness					
div. rate	multicellularity	−160.7763	0[a]	0.5453	*0.0007*
div. rate	sexuality	−156.3670	0[a]	0.4107	*0.0056*
div. rate	multicellularity (0.73) + sexuality (0.27)	−159.5671	0[a]	0.5660	*0.0029*
multicellularity	sexuality	164.6609	0[a]	0.5418	*0.0008*
ln-richness	div. rate	82.5635	0[a]	0.5384	*0.0008*
ln-richness	ln-age	95.0779	0[a]	0.0364	0.4636
projected richness (low)					
div. rate	multicellularity	−157.1000	0[a]	0.4905	*0.0017*
div. rate	sexuality	−154.2875	0[a]	0.3988	*0.0065*
div. rate	multicellularity (0.65) + sexuality (0.35)	−156.3698	0[a]	0.5272	*0.0053*
multicellularity	sexuality	163.4575	0[a]	0.5025	*0.0014*
ln-richness	div. rate	100.2158	0.826	0.5369	*0.0008*
ln-richness	ln-age	110.0502	0[a]	0.0115	0.6825
projected richness (high)					
div. rate	multicellularity	−156.0320	0[a]	0.5033	*0.0014*
div. rate	sexuality	−153.2296	0[a]	0.4143	*0.0053*
div. rate	multicellularity (0.64) + sexuality (0.36)	−155.4694	0[a]	0.5436	*0.0041*
multicellular	sexuality	163.2765	0[a]	0.5004	*0.0015*
ln-richness	div. rate	103.6561	0.86	0.5651	*0.0005*
ln-richness	ln-age	114.4356	0[a]	0.0164	0.6245

[a]A λ of 1 was significantly rejected ($p < 0.05$).

may be the monophyletic sister group to land plants [23,24]. Moreover, the estimated frequencies of multicellularity and sexual reproduction are very similar for Zygnemophyceae (30.5 and 100%) and Charophyta (38.3 and 99.5%; electronic supplementary material, datafiles S3 and S5). Therefore, treating this taxon in our analyses as Charophyta or Zygnemophyceae should have little impact on the results, because the phylogenetic position (sister to Embryophyta), age, richness, diversification rates and trait frequencies are similar either way. Furthermore, other subgroups of Charophyta could also have been used as terminal units in our analyses, given their phylogenetic distinctness from land plants and other algal clades [23,24] but the age and composition of these groups is still somewhat uncertain.

Besides the main analyses using 17 major clades, we also performed alternative analyses in which we treated bacteria as 14 distinct clades (phyla) and archaeans as two. This yielded 31 taxa overall, with a similar number of prokaryote and eukaryote clades ($n = 16$ and 15, respectively). We used previously compiled data on the phylogeny, species richness and age of these 16 prokaryotic phyla, and their relationships to other taxa in the tree [2]. Details are in the electronic supplementary material, appendix S3, and the tree for all 31 taxa is given in the electronic supplementary material, datafile S6. We were not able to project the undescribed richness among these clades. However, results were generally similar between the 31-clade and 17-clade analyses, suggesting that simply adding more bacterial species

would not overturn the results. We also performed analyses in which we included only the 15 eukaryotic clades, which yielded similar results. This allowed us to ensure that our main conclusions were not an artefact of comparing the (mostly) unicellular and asexual prokaryotes to the eukaryotes (which are variable for both traits). The tree for these 15 clades is given in the electronic supplementary material, datafile S7.

We acknowledge that we focused on a single estimate of phylogeny and divergence times for this study [2,21]. We know of few other estimates that are both time-calibrated and include all the relevant clades. Further, our phylogenetic regression analyses found little phylogenetic signal in the relationships between these variables (table 1), making these phylogenetic regression analyses equivalent to non-phylogenetic analyses. Thus, the details of the phylogeny should have little impact on the results. On the other hand, alternative phylogenies might influence the ages of clades, and thereby the estimated diversification rates. However, alternative trees should also show that land plants, animals and fungi are relatively young, and thus have high rates relative to other kingdom-level clades (given the high species richness of these three clades; figure 1), regardless of the details of the ages and phylogeny.

We note that we could have analysed the data at a lower taxonomic level. However, our primary interest in this study was in finding the traits that explain variation in diversification rates among the largest branches of the Tree of Life. Therefore,

we did not focus on explaining variation among subclades (e.g. within animals, plants or fungi). Of course, it would be interesting to analyse multicellularity and sexual reproduction within those clades that are variable for one or both traits (e.g. fungi). However, the results at lower taxonomic scales do not necessarily explain the large-scale patterns, and these patterns at the largest phylogenetic scales were our focus here. In other words, we chose the traits based on the phylogenetic scope: we did not choose the phylogenetic scope based on the traits.

(b) Multicellularity

After selecting the 17 focal clades, we estimated the frequencies of multicellularity and sexual reproduction among their species. The estimates are summarized in the electronic supplementary material, datafiles S5, S8 and S9, and supporting data and references are given in the electronic supplementary material, datafiles S2–S4.

We first estimated how many species in each clade were multicellular. We defined a multicellular organism as having cell-to-cell adherence and cell-to-cell communication [25]. In general, we considered a species to be multicellular if it was multicellular during at least part of its life cycle (given that most organisms which are considered multicellular are also unicellular during part of their life cycle, if only briefly). Some fungus species are dimorphic and have two different phenotypes: a mycelial, multicellular form and a yeast-like, unicellular form [26]. If these species had a multicellular, mycelial form, they were considered multicellular. Species characterized instead as exclusively colonial or yeast-like were considered unicellular, as the usual, dominant form in yeast is unicellular [27]. Colonial organisms have cell-to-cell adherence but not cytoplasmic (symplastic) continuity among adjoining cells [25]. Species in which individuals were generally unicellular but occasionally aggregated (but not as part of the regular life cycle) were not considered multicellular.

Clades reported to be entirely unicellular or multicellular were coded as such (electronic supplementary material, datafile S2). If a clade included both unicellular and multicellular taxa, we estimated the frequency of multicellularity among species, starting at the phylum level (electronic supplementary material, datafile S3). We searched the literature using Google Scholar, with the name of each higher taxon and 'unicellular' and then 'multicellular' as keywords. We used a list of phyla within each clade [2], but with four additional phyla in fungi (Chytridiomycota, Microspordia, Mucoromycota, Zygomucota). If a phylum included both unicellular and multicellular classes, we estimated the proportion of species with each state based on their frequency among classes. Similarly, we searched at the level of orders, families and genera when these taxa were variable. If a genus was not described as being multicellular or unicellular overall, we searched for data at the species level. We estimated the frequency of each state in that genus based on species with known states. For bacteria, which are predominantly unicellular, we performed a more limited search to estimate the frequency of multicellularity (details in the electronic supplementary material, appendix S4).

For each clade, the frequency of each state was estimated based on the proportion of species with each state in the higher taxa within the clade, and the richness of those taxa. Specifically, we multiplied the proportion by the richness of that higher taxon and then summed the richness for each state across the taxa within that clade. This was done at all taxonomic scales (e.g. phyla to genera) that were variable for this trait. Taxa with no data were excluded when calculating the proportion of multicellular species.

In general, we used species richness data summarized for major clades [2]. However, we also used the Catalogue of Life [28] (CoL), and other databases (see below) to estimate the richness of lower-level taxa. This sometimes led to slight differences

in richness for phyla and other higher taxa (relative to [2]), and we used these updated numbers instead.

For fungi, we used the CoL [28] for five phyla (Ascomycota, Basidiomycota, Glomeromycota, Microsporidia, Zygomycota). This database included more species of these phyla and the taxonomic composition of each phylum was relatively clear (e.g. in terms of family and genera). Taxonomic information above the family level was relatively clear in the MycoBank Database [29]. However, below the family level, the taxonomy was more complex (e.g. given different placements of species among genera). We used MycoBank [29] for three phyla (Blastocladiomycota, Chytridiomycota, Mucoromycota). Mucoromycota was not found in the CoL [28]. Blastocladiomycota was a class of Chytridiomycota in the CoL [28], not a separate phylum. We checked data carefully to avoid replicated genera. Subspecies, varieties and synonyms were not counted as species. Because two databases were used for the eight fungal phyla, we compared genera, families, orders and classes among these phyla in the two databases to avoid including the same taxon in different phyla.

Data were lacking in the CoL [28] for the algal phyla Bacillariophyta, Charophyta, Chlorophyta, Euglenozoa and Rhodophyta. We instead used AlgaeBase [30] for these taxa. AlgaeBase includes terrestrial, marine and freshwater algae, and is particularly complete for marine algae.

The phyla Apicomplexa, Dinophyta, Fornicata, Haptophytes and Rhizaria were not found in the CoL [28]. Furthermore, richness data for Amoebozoa were limited. Therefore, we used the National Center for Biotechnology Information (NCBI) taxonomy database [31] for these six phyla. Only species with formal scientific names in the NCBI taxonomy database were used to estimate species richness of phyla, classes, orders, families and genera. Entries based on environmental samples, varieties and unverified species were not included.

(c) Sexual reproduction

We also estimated the proportion of sexual and asexual species in each kingdom. Our definitions of asexual versus sexual reproduction followed Kondrashov [32]. Thus, we define asexual reproduction as reproduction in which each new individual originates from mitotic cell division, which does not change the genotype. Sexual reproduction is defined as the alteration of meiosis and syngamy with attendant segregation and recombination. Meiotic cell division involves halving the amount of DNA, genetic recombination from independent segregation of nonhomologous chromosomes and crossing over between homologous chromosomes. Syngamy involves the fusion of two haploid gametes. We give more detailed descriptions of different terms used to characterize reproduction in the electronic supplementary material, appendix S1.

Overall, we considered a species to be sexual if it was described as having sexual reproduction, or both sexual and asexual reproduction, or having a sexual stage. Only species described as being entirely asexual were considered asexual.

We calculated the proportion of sexual species in each kingdom following the general methodology described for multicellularity. However, for animals and land plants, large-scale summaries of trait frequencies were available. We describe how we estimated frequencies for these two clades in the electronic supplementary material, appendix S5. We also confirmed that similar overall results were obtained using a different frequency of sexual reproduction in land plants (electronic supplementary material, appendix S5).

For other clades, we searched the literature for data on reproduction in taxa within that clade. Specifically, we used the name of the taxon (e.g. phyla, classes, orders) and 'sexual' and then 'asexual' as keywords to search the literature using Google Scholar. However, there were fewer data available on sexuality than cellularity.

In cases in which we had to search for data at the genus level (i.e. for variable families), only the five largest genera in

the family were used when the family included more than five genera (the largest genera will be the most influential for determining the family-level frequency). If we could not find information for the five largest genera in the family, we included the sixth largest genus (or seventh, etc.), until we found data for at least five genera. For most fungal phyla, we searched for reproductive data at the species level.

Bacteria are not generally considered to have sexual reproduction, but they do have genetic exchange with other conspecific individuals. Therefore, we also tested if our results would be impacted by considering all bacteria to have sexual reproduction.

(d) Estimating diversification rates and species richness

We used a well-established approach for estimating diversification rates of higher-level taxa, the method-of-moments estimator for stem-group ages (MS estimator hereafter [19]). This approach only requires information on the age and species richness of each clade, rather than requiring a detailed time-calibrated phylogeny within each clade (which is needed for most alternative methods, and is lacking for many clades analysed here). Using the stem-group estimator, only one species must be included in the phylogeny per clade, and the approach can also be robust to incomplete sampling among clades [2]. This approach yields strong relationships between true and estimated rates in simulations, even when rates vary strongly between subclades [33] and when rates vary strongly within clades over time [34]. Therefore, it does not require constant rates within clades over time to be accurate, despite many assertions to the contrary. The approach also remains accurate when there are faster rates in younger clades [35], a pattern found to be widespread across the Tree of Life [2]. This approach does not attempt to disentangle the contribution of speciation and extinction rates to diversification rates, nor estimate variation in rates at different timepoints within a clade. We focus on explaining the current patterns of richness among these clades, even if richness within clades changed extensively over time.

The MS estimator uses a correction for clades that are entirely unsampled because they are extinct (ε, extinction fraction [19]). Therefore, a single value of ε is typically applied across all clades. Simulations show that the MS estimator is accurate when a single ε is assumed but speciation and extinction rates vary among clades [33,34]. Following standard practice, we used three values (0, 0.5, 0.9) but emphasize those from the intermediate value (0.5) in the main text. All three yielded similar relationships between traits and diversification rates. Again, simulations show that different values also yield similar relationships between true and estimated rates for the stem-group estimator, even when extinction rates vary from clade to clade [33,34].

We used three approaches for estimating the species richness of each clade for calculating diversification rates. First, we used the described species richness of each clade [2], as described above. We then used two alternative estimates [20] that incorporate projections of undescribed species (but focusing on specific clades projected to be particularly species rich). We describe these estimates in the electronic supplementary material, appendix S5. Note that these projections include bacteria, protists, fungi and animals.

Importantly, the use of these projections should account for potential bias if there is a greater propensity for researchers to describe multicellular rather than unicellular species. Specifically, most of the undescribed (projected) richness is among unicellular species of bacteria, protists and fungi. We think that using specific estimates of undescribed richness is the best way to deal with this potential bias. Note also that the projections used [20] were based on a review of all major groups across the Tree of Life (and only some of them were projected to have millions of undescribed species).

(e) Testing relationships between variables

We used phylogenetic generalized least-squares regression (PGLS) to test relationships between traits and diversification [36]. PGLS was implemented in the R package *caper* [37] v. 0.5.2. Following standard practice, we used the maximum-likelihood transformation of branch lengths, based on the estimated values of phylogenetic signal (λ) [38]. The use of λ estimates and corrects for the observed level of phylogenetic signal in the data. PGLS is valid when all variables are continuous, and when independent variables are categorical and the dependent variable is continuous [36], like diversification rates. Statistical significance was assessed using a sequential Bonferroni correction [39,40] for each table of regression results.

We tested for relationships between diversification (dependent variable) and multicellularity and sexual reproduction (independent variables). We tested each independent variable separately and then in combination. We then compared the AIC (Akaike information criterion [41]) for all three models to evaluate which had the best fit. Models within four AIC units of each other were not considered to have significantly different fit [41]. For the multiple regression model (including both multicellularity and sexual reproduction), we also calculated the standardized partial regression coefficients (using R code from [42]), to evaluate the contribution of each independent variable to the overall variance explained. The ability of this general approach to infer how much variance in diversification rates is explained by each variable (alone and in combination) is a major advantage relative to other approaches to studying diversification.

In addition to testing relationships between traits and diversification, we also used PGLS to evaluate how much variance in species richness among clades was explained by variation in diversification rates. A significant relationship between richness and diversification rates is not inevitable [2,35]. We also tested for a relationship between species richness and each clade's stem age, given that clades may be more species rich because they are older [2]. Finally, we tested for a relationship between multicellularity (independent variable) and sexual reproduction (dependent variable). Sexual reproduction might (hypothetically) drive multicellularity instead of the converse, but this should have little impact on the PGLS results.

Testing relationships between trait frequencies and diversification rates among higher-level clades has been done in many previous studies [42–45], including analyses across animals [43] and plants [44]. Nevertheless, there are scenarios whereby trait frequencies might appear to be related to diversification, but without a causal relationship. For example, a clade might consist of two subclades (A and B), with the trait of interest present only in subclade A but with increased diversification rates only in subclade B [43]. However, this problematic scenario would need to be repeated across multiple clades to generate a strong relationship between the trait and diversification rates, which seems unlikely. Note that the frequencies of traits within clades are not necessarily related to the number of trait origins. Thus, there could be multiple origins of a trait in a clade, but that trait could still be present at low frequencies among species (especially if the trait did not increase diversification rates). Alternatively, a trait could arise only once within a clade and be present in almost all the species, or may have arisen before the origin of that clade.

We used PGLS regression because this is a standard approach for analysing comparative data. However, the data were not normally distributed. We describe the normality tests and our non-parametric analyses in the electronic supplementary material, appendix S6, tables S7–S10.

3. Results

The estimated trait frequencies for each clade for multicellularity and sexual reproduction are given in the

electronic supplementary material, datafile S5, along with the estimated species richness, age and diversification rate of each clade. Data on trait frequencies, clade ages and richness are summarized in figure 1, along with the phylogeny. The baseline results are based on the number of described species in each clade and an intermediate extinction fraction ($\varepsilon = 0.5$) for estimating diversification rates (ε is the assumed ratio of extinction to speciation [19]). However, we also explored the robustness of the results to alternative extinction fractions, different projections of species richness and other assumptions.

These baseline results showed significant, positive relationships between multicellularity and diversification rates and between sexual reproduction and diversification (figure 2a,b and table 1). Multicellularity explained 54% of the variation in diversification rates among clades, whereas sexual reproduction explained 41%. A phylogenetic multiple regression model including both variables explained 57% of the variation in diversification rates among clades, but had slightly poorer fit than the one based on multicellularity alone (table 1). In the context of the model including both traits, 73% of the variance in diversification rates explained by the model was explained by multicellularity, and 27% by sexual reproduction. Variation in diversification rates also explained the majority of the variance in richness among these clades ($r^2 = 0.54$), but the ages of clades did not ($r^2 = 0.04$). There was a significant, positive relationship between the proportion of sexual and multicellular species among clades ($r^2 = 0.54$; figure 2c).

Results were generally similar using projected species numbers in major clades, rather than described species richness (table 1). Using a relatively low projected number of undescribed species (282 million species total, 77% bacteria; electronic supplementary material, datafile S8) showed more variance in diversification rates explained by multicellularity (49%) than sexuality (40%). A model including both variables again had poorer fit than the one based on multicellularity alone, and explained 53% of the variance in diversification rates. In this model, multicellularity again explained more variance than sexuality (65 versus 35%). The relationship between multicellularity and sexuality was also similar ($r^2 = 0.50$).

Results were also similar (table 1) using much larger projected species numbers (2.238 billion species total; 78% bacteria; electronic supplementary material, datafile S9). However, less variance in diversification rates was explained by multicellularity alone (50%). A model including both multicellularity and sexuality explained 54% of the variance in diversification rates (but again with poorer fit than the model including multicellularity alone). Multicellularity again explained more variance than sexuality (64 versus 36%) in this two-trait model. Note that in both of these analyses using projected richness, we included large projected numbers for bacteria, protists, fungi, and animals, and not only bacteria.

These general results were robust to changing other assumptions, beyond species richness. There was very little effect of using alternative extinction fractions ($\varepsilon = 0$ and 0.9) to estimate diversification rates, for all three sets of species numbers (electronic supplementary material, tables S1 and S2). Assuming that bacteria all have sexual reproduction also had relatively minor effects (electronic supplementary material, table S3). Overall, the variance in diversification rates explained by sexual reproduction declined slightly when bacteria were considered to have sexual reproduction.
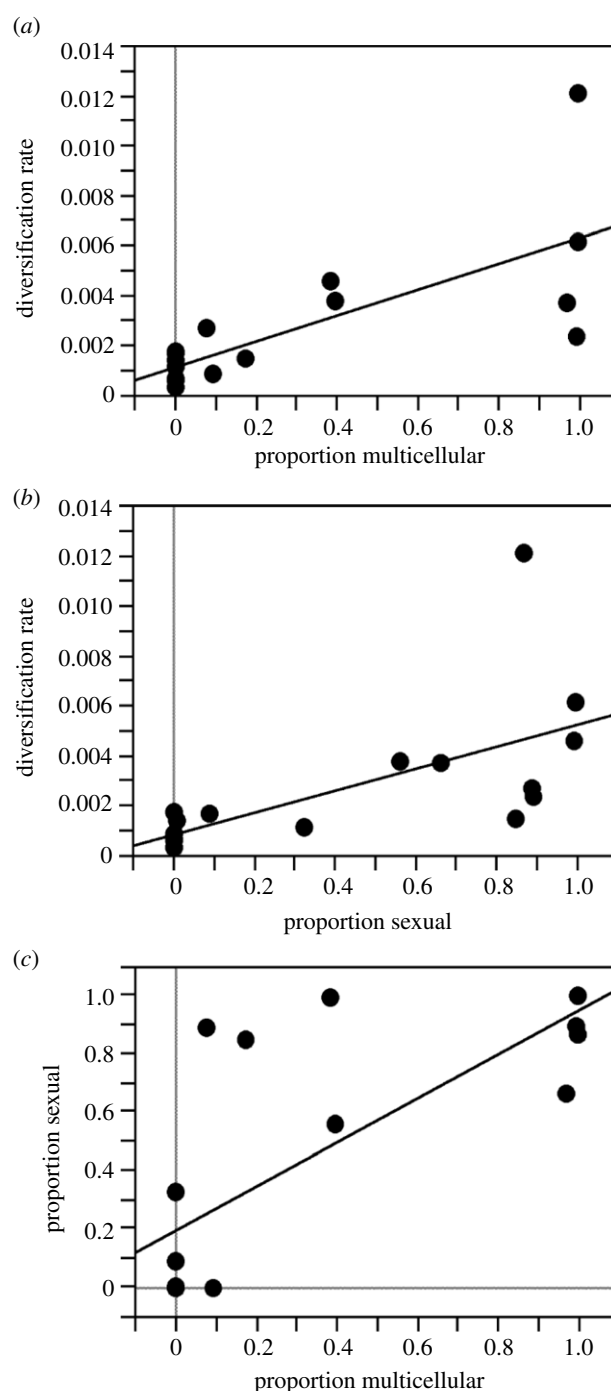

Figure 2. Plots showing relationships between (a) estimated diversification rates of the 17 clades and their estimated proportions of multicellular species, (b) estimated diversification rates and their proportions of species with sexual reproduction, and (c) the estimated proportions of sexual and multicellular species. Results are based on described species richness and diversification rates estimated using $\varepsilon = 0.5$. Results are similar using projected richness (table 1) and alternative ε values (electronic supplementary material, tables S2 and S3). For the ease of interpretation, results based on raw values are shown. The main results (table 1) are based on phylogenetic regression, but are almost identical.

We also performed analyses in which we treated prokaryotes as 16 separate clades (i.e. phyla) instead of two (archaeans, bacteria), yielding 31 clades in total and similar numbers of prokaryotic and eukaryotic clades (electronic supplementary material, datafile S10). The baseline results ($\varepsilon = 0.5$, described richness; electronic supplementary material, table S4) were similar to those using 17 clades

(table 1), but the impact of multicellularity was weakened. We found significant relationships between multicellularity and diversification ($r^2 = 0.40$) and sexuality ($r^2 = 0.52$). A model with both variables explained more variance (58%) and had the best fit. We also included the eukaryotic cell as a separate trait in these analyses (electronic supplementary material, table S4). This trait was significantly related to diversification rates ($r^2 = 0.24$), but an analysis with all three variables had poorer fit than the one with just multicellularity and sexuality. When we treated all bacteria as having sexual reproduction in this analysis (electronic supplementary material, table S5), the effect of sexuality on diversification rates was weaker ($r^2 = 0.18$), and the best-fitting model included multicellularity alone. These alternative analyses with 31 clades also showed strong, positive relationships between species richness and diversification rates ($r^2 = 0.61$) but not clade ages ($r^2 = 0.23$; negative relationship).

Results were also similar to the main results (table 1) when analysing eukaryotes alone (15 clades; electronic supplementary material, table S6), with a strong relationship between multicellularity and diversification ($r^2 = 0.52$), a weaker relationship with sexuality ($r^2 = 0.37$) and a combined model that was dominated by multicellularity and had poorer fit than the one including multicellularity alone. Again, there were strong relationships between species richness and diversification rates ($r^2 = 0.59$) but not clade ages ($r^2 = 0.11$).

Relationships between multicellularity and diversification were broadly similar among the main analyses ($r^2 = 0.50$–0.54; table 1) and those using 31 clades ($r^2 = 0.40$; electronic supplementary material, table S4) or only the 15 eukaryotic clades ($r^2 = 0.52$; electronic supplementary material, table S6). The relationship between sexuality and multicellularity was weaker using 31 clades and treating all bacteria as sexual ($r^2 = 0.29$; electronic supplementary material, table S5), but similar to the 17-clade analyses making this assumption ($r^2 = 0.42$; electronic supplementary material, table S3).

Finally, we performed a non-parametric version of our main analyses. These analyses yielded similar results to those from PGLS, supporting the influence of multicellularity (strongly) and sexuality (more weakly) on diversification rates, and the strong correlation between multicellularity and sexuality (electronic supplementary material, appendix S6).

4. Discussion

In this study, we provide evidence that multicellularity and sexual reproduction both strongly influenced diversification rates of the major clades across the Tree of Life. Together, these two variables statistically explain more than half of the variance in diversification rates among these clades. Variation in diversification rates then explains most variation in species richness among these clades. We also find that multicellularity is generally more important than sexuality in explaining these diversification patterns, and that multicellularity and sexual reproduction are strongly related to each other in their distribution among clades.

Our results offer possibly the first analysis of the traits that may explain diversity patterns across the entire Tree of Life, and should open the door for future studies that address the specific mechanisms by which these traits might increase speciation and/or reduce extinction within clades. For example, multicellularity allows for differentiation into many different tissue types, which might increase organismal complexity and thereby possibly accelerate rates of divergence among species within clades [6,7].

We acknowledge that our results are based on a statistical relationship between these variables, and so causation is not proven. This will be the case for almost any empirical macro-evolutionary study, regardless of trait, taxa or scale. Similarly, it is possible that other traits explain these diversification patterns instead of multicellularity or sexual reproduction. However, it is unclear what these traits would be, especially given the lack of obvious commonalities uniquely shared among the fastest diversifying clades (animals, plants, fungi). It is also possible that there are simply unique traits within each of these clades that explain their rapid diversification. Yet, we suspect that many potential candidate traits would be contingent on multicellularity and/or sexual reproduction (e.g. flowers in plants).

Our findings may be relevant to explaining the paradox of sex [9–13]. This paradox is usually stated as: why is sexual reproduction so prevalent among species despite its apparent disadvantages? Much literature on this topic focuses on empirical and theoretical investigations of closely related species and populations (typically multicellular organisms) and why sex is maintained over time [13]. Our results suggest that sexual reproduction may also be numerically widespread among species (at least in part) because it increased rates of diversification among major clades at deep timescales.

How might sexual reproduction increase diversification? Previous theoretical research has suggested that asexual lineages may have higher extinction rates than sexual lineages, potentially caused by the accumulation of deleterious mutations (Muller's ratchet [46,47]) or a lack of genetic diversity that hinders their evolutionary response to biotic and/or abiotic threats [10], including parasites [11]. Thus, secondarily asexual eukaryotes are thought to be 'evolutionary dead ends', although the evidence is somewhat mixed [16]. Sexual reproduction might also increase speciation rates, since it may increase overall rates of adaptation and evolution [9–11]. A simulation study found that sexual reproduction increased speciation rates (relative to asexual lineages) but did not lead to higher richness of sexual species [48].

Our results raise another possible explanation for the prevalence of (described) sexual species. Sexual reproduction may be widespread (at least in part) because sexual reproduction is associated with multicellularity, and multicellularity is instead the main driver of these large-scale patterns of diversification and richness among major clades.

We also note that sexual reproduction is only predominant based on described species numbers: some projections suggest that approximately 78% of all species are bacteria [20]. If these projections are accurate, then species with sexual reproduction represent only a minority of all species. Indeed, the paradox of species is specifically that the majority of species have sexual reproduction, despite its apparent disadvantages [12].

We found strong relationships between multicellularity and sexual reproduction among these major clades. Our results do not resolve which trait drives the evolution of the other. Nevertheless, they do imply a possible causal relationship between these traits at broad phylogenetic scales. Some previous hypotheses and theory have proposed that multicellularity precedes and drives the evolution of differentiated male and female sexes, starting with differentiated gametes

[11,17]. There is some empirical support for this hypothesis within volvocine algae [18]. Further testing of the causes of the observed relationship between sexuality and multicellularity among these major clades will be an important area for future research.

Many other questions are raised by these results. Our analyses explain roughly 50% of the variance in diversification rates among these clades, but considerable variance remains unexplained. What explains the rest? Some of the unexplained variance may be attributed to the exceptionally high diversification rate in land plants (Embryophyta), which is almost certainly related to the very fast diversification rate in angiosperms [44]. This very high rate in land plants (coupled with lower species richness than in animals) may also weaken the overall relationship between diversification rates and species richness among these clades (table 1). Considerable unexplained variance in diversification rates might also result from groups that have multicellularity and/or sexual reproduction but unexceptional diversification rates. For example, the marine algae clade Rhodophyta has both traits at high frequencies but modest diversification rates. Why has this group failed to radiate as rapidly as other clades with these traits? Studies in animals [43] suggest that predominantly marine clades have lower diversification rates than mostly terrestrial ones (and land plants, animals and fungi are predominantly terrestrial). A similar explanation may apply to largely freshwater algae groups (Charophyta, Chlorophyta), given results in vertebrates showing lower rates in aquatic groups in general, not just marine groups [45]. An extensive dataset on the habitats of these clades will be needed to test these patterns. We also found mostly unicellular clades which have sexual reproduction at high frequencies but relatively low diversification rates (Amoebozoa, SAR clade). These groups may help explain the weaker relationships between diversification and sexual reproduction, relative to multicellularity.

Finally, we recognize that readers may have a diversity of valid concerns about the methods of our study. In addition to the Material and methods section, we address these at length in the electronic supplementary material, appendix S7. These involve potential errors in estimating trait frequencies, variability in reproductive modes, whether bacteria have sexual reproduction, the problem of comparing species across all of life, and concerns about the diversification-rate estimators. We urge readers that are concerned about these issues to consult the electronic supplementary material, appendix S7.

5. Conclusion

We found that much of the variation in diversification rates (and species richness) among the major branches of the Tree of Life is related to the positive effects of multicellularity and sexual reproduction on diversification rates. These results help explain why three disparate groups of organisms (animals, land plants, fungi) have been so evolutionarily successful in terms of species numbers. Moreover, our results may have implications for the 'paradox of sex' (i.e. the dominance of sexual reproduction among species, despite its disadvantages). Sexual reproduction may be widespread relative to asexual reproduction (in part) because it increases diversification rates among major clades, and possibly because of its association with multicellularity (given that multicellularity has a stronger impact on diversification rates). We show that these overall results are robust regardless of whether overall species richness on Earth is in the millions or billions.

References

1. Futuyma DJ. 2009 *Evolution*. Sunderland, MA: Sinauer Associates.

2. Scholl JP, Wiens JJ. 2016 Diversification rates and species richness across the Tree of Life. *Proc. R. Soc. B* **283**, 20161335. (doi:10.1098/rspb.2016.1334)

3. Ricklefs RE. 2007 Estimating diversification rates from phylogenetic information. *Trends Ecol. Evol.* **22**, 601–610. (doi:10.1016/j.tree.2007.06.013)

4. Morlon H. 2014 Phylogenetic approaches for studying diversification. *Ecol. Lett.* **17**, 508–525. (doi:10.1111/ele.12251)

5. Maynard SJ, Szathmary E. 1995 *The major transitions in evolution*. Oxford, UK: Freeman.

6. Carroll SB. 2001 Chance and necessity: the evolution of morphological complexity and diversity. *Nature* **409**, 1102–1109. (doi:10.1038/35059227)

7. Grosberg RK, Strathmann RR. 2007 The evolution of multicellularity: a minor major transition? *Annu. Rev. Ecol. Evol. Syst.* **38**, 621–654. (doi:10.1146/annurev.ecolsys.36.102403.114735)

8. Schirrmeister B, de Vos JM, Antonelli A, Bagheri HC. 2013 Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. *Proc. Natl Acad. Sci. USA* **110**, 1791–1796. (doi:10.1073/pnas.1209927110)

9. Williams GC. 1975 *Sex and evolution*. Princeton, NJ: Princeton University Press.

10. Maynard Smith J. 1978 *The evolution of sex*. New York, NY: Cambridge University Press.

11. Bell G. 1982 *The masterpiece of nature. The evolution and genetics of sexuality*. Berkeley, CA: University of California Press.

12. Otto SP, Lenormand T. 2002 Resolving the paradox of sex and recombination. *Nat. Rev. Genet.* **3**, 252–261. (doi:10.1038/nrg761)

13. Neiman M, Lively CM, Meirmans S. 2017 Why sex? A pluralist approach revisited. *Trends Ecol. Evol.* **32**, 589–600. (doi:10.1016/j.tree.2017.05.004)

14. Fontaneto D, Tang CQ, Obertegger U, Leasi F, Barraclough TG. 2012 Different diversification rates between sexual and asexual organisms. *Evol. Biol.* **39**, 262–270. (doi:10.1007/s11692-012-9161-z)

15. Judson OP, Normark BB. 1996 Ancient asexual scandals. *Trends Ecol. Evol.* **11**, 41–46. (doi:10.1016/0169-5347(96)81040-8)

16. Schwander T, Crespi BJ. 2009 Twigs on the tree of life? Neutral and selective models for integrating macroevolutionary patterns with microevolutionary processes in the analysis of asexuality. *Mol. Ecol.* **18**, 28–42. (doi:10.1111/j.1365-294X.2008.03992.x)

17. Bulmer MG, Parker GA. 2002 The evolution of anisogamy: a game-theoretic approach. *Proc. R. Soc. B* **269**, 2381–2388. (doi:10.1098/rspb.2002.2161)

18. Hanschen ER *et al.* 2018 Multicellularity drives the evolution of sexual traits. *Am. Nat.* **192**, E93–E105. (doi:10.1086/698301)

19. Magallon S, Sanderson MJ. 2001 Absolute diversification rates in angiosperm clades. *Evolution* **55**, 1762–1780. (doi:10.1111/j.0014-3820.2001.tb00826.x)

20. Larsen BB, Miller EC, Rhodes MK, Wiens JJ. 2017 Inordinate fondness multiplied and redistributed: the number of species on Earth and the new Pie of Life. *Q. Rev. Biol.* **92**, 229–265. (doi:10.1086/693564)

21. Parfrey LW, Lahr DJG, Knoll AH, Katz LA. 2011 Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl Acad. Sci. USA* **108**, 13 624–13 629. (doi:10.1073/pnas.1110633108)

22. Kumar S, Stecher G, Suleski M, Hedges SB. 2017 TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819. (doi:10.1093/molbev/msx116)

23. Wickett NJ *et al.* 2014 Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl Acad. Sci. USA* **111**, E4859–E4868. (doi:10.1073/pnas.1323926111)

24. Leebens-Mack J *et al.* 2019 One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685. (doi:10.1038/s41586-019-1693-2)

25. Niklas KJ, Newman SA. 2013 The origins of multicellular organisms. *Evol. Dev.* **15**, 41–52. (doi:10.1111/ede.12013)

26. Pfeiffer T, Schuster S, Bonhoeffer S. 2001 Cooperation and competition in the evolution of ATP-producing pathways. *Science* **292**, 504–507. (doi:10.1126/science.1058079)

27. Scherr GH, Weaver RH. 1953 The dimorphism phenomenon in yeasts. *Bacteriol. Rev.* **17**, 51–92. (doi:10.1128/br.17.1.51-92.1953)

28. Roskov Y *et al.* 2019 Species 2000 & ITIS Catalogue of Life, 2019 annual checklist. See www.catalogueoflife.org/annual-checklist/2019.

29. Crous PW, Gams W, Stalpers JA, Robert V, Stegehuis G. 2004 MycoBank: an online initiative to launch mycology into the 21st century. *Stud. Mycol.* **50**, 19–22.

30. Guiry MD, Guiry GM. 2019 Algaebase. Galway: world-wide electronic publication. National University of Ireland. See http://www.algaebase.org.

31. Sayers EW *et al.* 2009 Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **37**, D5–D15. (doi:10.1093/nar/gkn741)

32. Kondrashov AS. 1988 Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440. (doi:10.1038/336435a0)

33. Meyer ALS, Wiens JJ. 2018 Estimating diversification rates for higher taxa: BAMM can give problematic estimates of rates and rate shifts. *Evolution* **72**, 39–53. (doi:10.1111/evo.13378)

34. Meyer ALS, Román-Palacios C, Wiens JJ. 2018 BAMM gives misleading rate estimates in simulated and empirical datasets. *Evolution* **72**, 2257–2266. (doi:10.1111/evo.13574)

35. Kozak KH, Wiens JJ. 2016 Testing the relationships between diversification, species richness, and trait evolution. *Syst. Biol.* **65**, 975–988. (doi:10.1093/sysbio/syw029)

36. Martins EP, Hansen TF. 1997 Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* **149**, 646–667. (doi:10.1086/286013)

37. Orme DL. 2013 The caper package: comparative analysis of phylogenetics and evolution in R. R package version 0.5.2. See https://cran.r-project.org/web/packages/caper/index.html.

38. Pagel M. 1999 Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884. (doi:10.1038/44766)

39. Holm S. 1979 A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70. (doi:10.2307/4615733)

40. Rice WR. 1989 Analyzing tables of statistical tests. *Evolution* **43**, 223–225. (doi:10.1111/j.1558-5646.1989.tb04220.x)

41. Burnham KP, Anderson DR. 2002 *Model selection and multi-model inference: a practical information-theoretic approach*. New York, NY: Springer Press.

42. Moen DS, Wiens JJ. 2017 Microhabitat and climatic-niche change explain patterns of diversification among frog families. *Am. Nat.* **190**, 29–44. (doi:10.1086/692065)

43. Jezkova T, Wiens JJ. 2017 What explains patterns of diversification and richness among animal phyla? *Am. Nat.* **189**, 201–212. (doi:10.1086/690194)

44. Hernández-Hernández T, Wiens JJ. 2020 Why are there so many flowering plants? A multi-scale analysis of plant diversification. *Am. Nat.* **195**, 948–963. (doi:10.1086/708273)

45. Wiens JJ. 2015 Explaining large-scale patterns of vertebrate diversity. *Biol. Lett.* **11**, 20150506. (doi:10.1098/rsbl.2015.0506)

46. Muller H. 1964 The relation of recombination to mutation advance. *Mutat. Res.* **1**, 2–9. (doi:10.1016/0027-5107(64)90047-8)

47. Lynch M, Burger R, Butcher D, Gabriel W. 1993 The mutational meltdown in asexual populations. *J. Hered.* **84**, 339–344. (doi:10.1093/oxfordjournals.jhered.a111354)

48. Melián CJ, Alonso D, Allesina S, Condit RS, Etienne RS. 2012 Does sex speed up evolutionary rate and increase biodiversity? *PLoS Comput. Biol.* **8**, e1002414. (doi:10.1371/journal.pcbi.1002414)

49. Chen L, Wiens JJ. 2021 Data from: Multicellularity and sex helped shape the Tree of life. Dryad Digital Repository. (https://doi.org/10.5061/dryad.866t1g1r1)