



HAL
open science

Macroecological and macroevolutionary patterns emerge in the universe of GNU/Linux operating systems

Petr Keil, A. A. M. Macdonald, Kelly S. Ramirez, Joanne M. Bennett,
Gabriel E. Garcia-Pena, Benjamin Yguel, B erenger Bourgeois, Carsten Meyer

► To cite this version:

Petr Keil, A. A. M. Macdonald, Kelly S. Ramirez, Joanne M. Bennett, Gabriel E. Garcia-Pena, et al.. Macroecological and macroevolutionary patterns emerge in the universe of GNU/Linux operating systems. *Ecography*, 2018, 41 (11), pp.1788-1800. 10.1111/ecog.03424 . hal-02621181

HAL Id: hal-02621181

<https://hal.inrae.fr/hal-02621181>

Submitted on 26 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche franais ou  trangers, des laboratoires publics ou priv es.



Distributed under a Creative Commons Attribution 4.0 International License

ECOGRAPHY

Research

Macroecological and macroevolutionary patterns emerge in the universe of GNU/Linux operating systems

Petr Keil, A. A. M. MacDonald, Kelly S. Ramirez, Joanne M. Bennett, Gabriel E. García-Peña, Benjamin Yguel, Béranger Bourgeois and Carsten Meyer

P. Keil (<http://orcid.org/0000-0003-3017-1858>) (pkeil@seznam.cz) and *J. M. Bennett*, German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany. *JMB* also at: Inst. of Biology, Martin Luther Univ. Halle-Wittenberg, Halle (Saale), Germany. – *A. A. M. MacDonald*, *G. E. García-Peña* and *B. Bourgeois*, Centre de Synthèse et d'Analyse sur la Biodiversité – CESAB, Aix-en-Provence, France. *AAMM* also at: Univ. of British Columbia, Vancouver, BC, Canada. *GEG-P* also at: Centro de Ciencias de la Complejidad (C3) y Facultad de Medicina Veterinaria y Zootecnia, México. *BB* also at: Agroécologie – Agrosup Dijon, Inst. national de la recherche agronomique (INRA), Univ. Bourgogne Franche-Comté, Dijon, France. – *K. S. Ramirez*, Terrestrial Ecology, Netherlands Inst. of Ecology, Wageningen, the Netherlands. – *B. Yguel*, Unité MECADEV mécanismes adaptatifs et évolution, UMR 7179 CNRS/MNHN, Brunoy, France, and Centre d'Ecologie et des Sciences de la Conservation (CESCO-UMR 7204), Sorbonne Univ.-MNHN-CNRS-UPMC, CP51, Paris, France. – *C. Meyer*, Macroecology and Society, German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany, and Faculty of Biosciences, Pharmacy and Psychology, Univ. of Leipzig, Leipzig, Germany, and Dept of Ecology and Evolutionary Biology, Yale Univ., New Haven, CT, USA.

Ecography

41: 1788–1800, 2018

doi: 10.1111/ecog.03424

Subject Editor: Thiago Rangel
Editor-in-Chief: Miguel Araújo
Accepted 2 February 2018



What leads to classically recognized patterns of biodiversity remains an open and contested question. It remains unknown if observed patterns are generated by biological or non-biological mechanisms, or if we should expect the patterns to emerge in non-biological systems. Here, we employ analogies between GNU/Linux operating systems (distros), a non-biological system, and biodiversity, and we look for a number of well-established ecological and evolutionary patterns in the Linux universe. We demonstrate that patterns of the Linux universe generally match macroecological patterns. Particularly, Linux distro commonness and rarity follow a skewed distribution with a clear excess of rare distros, we observed a power law mean-variance scaling of temporal fluctuation, but there is only a weak relationship between niche breadth (number of software packages) and commonness. The diversity in the Linux universe also follows general macroevolutionary patterns: the number of phylogenetic lineages increases linearly through time, with clear per-species diversification and extinction slowdowns, something that has been indirectly estimated, but not directly observed in biology. Moreover, the composition of functional traits (software packages) exhibits significant phylogenetic signal. The emergence of macroecological patterns across Linux suggests that the patterns are produced independently of system identity, which points to the possibility of non-biological drivers of fundamental biodiversity patterns. At the same time, our study provides a step towards using Linux as a model system for exploring macroecological and macroevolutionary patterns.

Keywords: complex systems, cultural evolution, Debian, speciation, species-abundance distribution, Taylor's power law



www.ecography.org

Introduction

Despite the general paucity of strict laws in ecology and evolution (Lawton 1999), there are several quantitative patterns that are consistently observed across taxonomic, geographical and temporal scales. For example, abundances of species generally have skewed frequency distributions, with many more rare species than abundant ones (Gaston 1996a, McGill et al. 2007, Morlon et al. 2009), and there is a power-law relationship between the mean and the variance of temporal abundance fluctuations (Taylor and Woiwod 1980, Kendal 2004). In general, species diversity decreases from the equator towards the poles (Rosenzweig 1995, Gaston 2000) and increases with area (Rosenzweig 1995, Drakare et al. 2006). Species with large geographic distributions tend to have wider niches, and vice versa (Slatyer et al. 2013). Species with similar functional traits are often phylogenetically related (Darwin 1859, Losos 2008). Diversification of evolutionary lineages slows down over evolutionary time (Rabosky and Lovette 2008, Rabosky and Glor 2010, Moen and Morlon 2014), although generality of the latter pattern has been questioned (Harmon and Harrison 2015, Graham et al. 2016).

To explain these patterns, we can invoke uniquely ecological and evolutionary processes: the patterns could be an outcome of assembly rules, natural selection, behavior, species interactions, or interplay between specific functional traits and environments. However, it has been demonstrated that some of the patterns are not unique to ecological and evolutionary systems, and often emerge in other complex systems (Gaston et al. 1993, Mace and Pagel 1995, Bettencourt et al. 2007, Nekola and Brown 2007, Warren et al. 2011, Blonder et al. 2014, Scheffer et al. 2017). Examples are: species-abundance distributions of music festival setlists (Nekola and Brown 2007) and basketball wins by a team (Warren et al. 2011), frequency distributions of components of software (Pang and Maslov 2013), latitudinal gradients of language diversity (Mace and Pagel 1995), species–area relationships in corporations, industrial codes, and minerals (Blonder et al. 2014), and diversification slowdowns in American automobiles (Gjesfeld et al. 2016). To explain such conspicuous universality, theories have been proposed that predict the patterns across different systems for pure statistical reasons, given that the systems share structural constraints (Frank 2009, Sizling et al. 2009, Harte 2011, Blonder et al. 2014). One such typical constraint is the partition of populations of objects into categories.

However, structural constraints are not the only way that biological and non-biological systems can resemble one another – similarity may also arise from analogous underlying processes. Striking examples can be found in the rapidly expanding literature on cultural evolution (Dawkins 1976, Cavalli-Sforza and Feldman 1981, Boyd and Richerson 1985, Sperber 1996, Mesoudi 2016) which exploits parallels between (mostly) Darwinian evolution and the evolution of languages, beliefs, skills, knowledge, institutions or other forms of socially transmitted information. There has

also been a continuous feedback between evolutionary biology, ecology and economics (Malthus 1798, Maynard Smith 1982, Worster 1994, Bonds et al. 2012), and some adoption of ecological and evolutionary principles for understanding the development and maintenance of software (Fortuna et al. 2011, Valverde and Sole 2015). However, these fields have been emphasizing micro-evolutionary, population-genetic, or population-dynamic processes, lacking an evaluation of the emerging macro-ecological and macro-evolutionary patterns. The notable exceptions (Mace and Pagel 1995, Nekola and Brown 2007, Blonder et al. 2014, Gjesfeld et al. 2016) often focus on one selected pattern or on multiple but closely related ecological patterns (Blonder et al. 2014). While there are many examples of non-biological systems following patterns observed in biology, we are unaware of a system simultaneously exhibiting macroecological and macroevolutionary patterns.

Here, we propose structural analogies between biological diversity and the universe of open source GNU/Linux-based computer operating systems (hereafter Linux). We have chosen Linux, since its components are all openly available and well-documented, and because a large number of structural analogies can be drawn with the biological system of evolution, distribution and abundance of biological species, hereafter biodiversity (see Table 1 and Material and methods for details). Based on these analogies, we test the Linux universe for classical macroecological and macroevolutionary patterns and relationships, and we discuss the structural properties that likely generate common patterns in both biological and computer operating systems.

Material and methods

Biodiversity structures in Linux

The development and distribution of GNU/Linux operating systems uses various open-source licenses (e.g. GNU GPL2, <www.gnu.org>), which allows the code to be freely copied and modified. This has led to the development of hundreds of operating systems, connected in a branching pattern of descent with modification. These different operating systems possess different applications, and are used with various degrees of popularity. We propose that each Linux-based operating system, commonly referred to as a distro (Table 1) can be viewed as a lineage or species (Mens et al. 2014a, b). Our decision to give a Linux distro a distinct name, rather than calling it another version of an existing distro, was based on developers' subjective view that the distro is sufficiently different, which is similar to how many biological species have been defined to date.

The number of hardware devices on which a distro is installed can be interpreted as its commonness (e.g. population abundance or related, but not identical, range size). While popular distros can spread across devices, unpopular ones may go extinct. Thus, because distro abundances change

Table 1. Summary of analogies between ecological terms and the ‘Linux universe’, i.e. the system encompassing all Linux distros, Linux software packages, their developers and users, and relationships between those.

Ecological terminology	Linux definition
Species or lineage	A Linux distribution, most commonly referred to as ‘distro’. A distro is a computer operating system comprised of a collection of software packages.
Phylogeny	Genealogical relationships between distros. Most Linux distros have evolved from one of three main distros: Debian, Red Hat or SLS. We are not aware of a merging between distros, and hence a tree seems to be a good representation of Linux evolutionary history.
Commonness (abundance, range size)	Hits per day (HPD) or number of machines that run a distro. HPD is a yearly average of the number of times per day any given distro page on DistroWatch.com is accessed. HPD is a proxy for distro popularity.
Diversification event	Approximate date at which the development of a distro began.
Extinction event	Approximate date at which development on a distro ceased. De-extinction may occasionally happen, if development of a distro is resumed.
Functional or life-history traits	Software packages available with each distro. These packages or ‘functional traits’ determine the applicability of the distro.
Natural selection	Use and popularity of a distro is based on users and downloads. Unused and unpopular distros go extinct.
Niche breadth	Number of functions, purposes, or capabilities that a distro has. Similar to multidimensional niche volume.
Area/Productivity	Population of country where distro was developed. Alternatively, this could be the user base of each distro, however those statistics are not available.

through time, distros have population dynamics. Further, computer operating systems, and especially Linux distros, come with diverse functionalities, which are pre-installed in the form of software packages, and which can be made analogous to functional traits of biological species. Finally, new distros emerge through a process similar to biological speciation: When developers decide to come up with a new distro, part of the code of an ancestor distro is reused and combined with custom-tailored new code, and with existing open-source packages. The evolution of distros is influenced by environmental factors such as hardware architecture and user requirements (Yan et al. 2010); and constrained by user habits and the need for cost-effective development through reuse of code (Myers 2003, Fortuna et al. 2011); as a result, Linux distros have a genealogy which is potentially analogous to a phylogenetic tree.

Once the qualitative structural analogies were set, we examined if the analogous structural elements of Linux follow the same quantitative patterns as biodiversity. During this process, we found that some biodiversity patterns were hard to imagine in the Linux realm; such as the latitudinal gradient of biodiversity, which would depend on a reasonable analogy to latitude. Hence, from the set of known biodiversity patterns, we chose those for which reasonable analogies can be made given available data on Linux. Specifically, we looked for the following macroecological patterns: 1) the skewed species-abundance distributions, 2) power-law mean-variance scaling of population fluctuations, and 3) a positive relationship between niche breadth and range size. Between macroevolutionary patterns we looked for: 4) slowdowns of diversification rates over evolutionary time and 5) a phylogenetic signal in functional traits.

We note that the analogies between Linux and biology (Table 1) are imperfect. For example, the evolution of Linux distros involves no genetic inheritance, there is little individual-level variation, and the inheritance is not from one

user’s computer to another, but is instead enforced centrally by developers. Thus, the micro-evolutionary mechanism is not Darwinian *sensu stricto* (Lewontin 1970). This is a problem that is widely discussed in the field of cultural evolution (Mesoudi 2017), and it is usually concluded that even imperfect analogies are useful. In our case the analogies work well at the macro-evolutionary level – there certainly is a selection by users, there is origination and extinction, and distros do change in time. As a result, a clear phylogeny emerges, with each tip of the phylogeny (distro) having a unique and inherited combination of traits. We can then argue that if biodiversity and Linux share similar categorization of objects, but they differ in their inner processes and mechanisms, and yet they display the same emerging patterns, then we have a case for a non-biological and non-mechanistic explanation of these patterns.

Data on species commonness

We considered popularity (i.e. the number of users) of each distro as a measure of species commonness, which is analogous to either species abundance or range size. However, because most GNU/Linux operating systems are freely available for download and use, there is no single way to directly determine how frequently each distro is used. Therefore, we used three sources of data as proxies of commonness.

1) We used popularity metrics from Distrowatch (<www.distrowatch.com>), measured as hits per day (HPD), which is the daily number of clicks on each distro-specific page on Distrowatch. We assume that HPD correlated to the actual number of users, i.e. to the number of individuals in the global population of a particular Linux species. Although this assumption is problematic (Edge 2011), we still consider the HPD worth exploring, but we also consider alternative measures. From Distrowatch we extracted data on HPD of 275 distros averaged over a yearly period between 27/04/2016

and 28/04/2015 taken from <http://goo.gl/hMjUXr>. These data were used to examine the species-abundance distributions (SAD), and to assess the relationship between niche breadth and commonness. We also extracted HPD for each year between (and including) 2002 to 2015 (from the main page on www.distrowatch.com, section Page Hit Ranking; downloaded on 28/04/2015). In each year, these data were available only for the 100 most popular distros. This makes them unsuitable for any comprehensive temporal dynamics of SAD, but it allows analyses of population dynamics for the more common distros. We used distros that were present in the data for ≥ 10 consecutive years to assess the temporal mean-variance scaling (Taylor's power law; TPL).

2) As an alternative and a more direct measure of distro commonness we used the sample of 164,726 computers registered, by volunteers, at LinuxCounter (www.linuxcounter.net/statistics/distributions); downloaded on 27/11/2017), where the identity of each distro on each computer is known. These data were used to examine the species-abundance distributions (SAD), and to assess the relationship between niche breadth and commonness.

3) Finally, we used Wikimedia traffic analysis reports (<http://goo.gl/2h6Rq6>), which give monthly counts

of requests (squids) on Wikipedia for pages on specific operating systems. We downloaded two sets of data: 1) data logged between 2010 and 2011 for 8 most popular distros at the time, and 2) data logged between 2012 and 2014 for 17 most popular distros at the time. These were used for calculations of Taylor's Power Law (TPL), separately in each of the two time periods.

Models of species-abundance distribution (SAD)

We fitted the lognormal and logseries models to the data on frequency of commonness from Distrowatch (HPD) and LinuxCounter (no. of machines). We chose these models since they describe well the generally observed excess or rare species in ecological abundance data (Baldrige et al. 2016). We fit the logseries model using package `sads` in R (function `fits`). To fit the lognormal model, we used the mean and standard deviation of log commonness as the maximum likelihood estimates of parameters of the lognormal probability density function. To plot the fitted models alongside the data, we drew a random sample from each distribution and ordered the outcomes. This was repeated 500 times and the average is shown as the solid lines in Fig. 1.

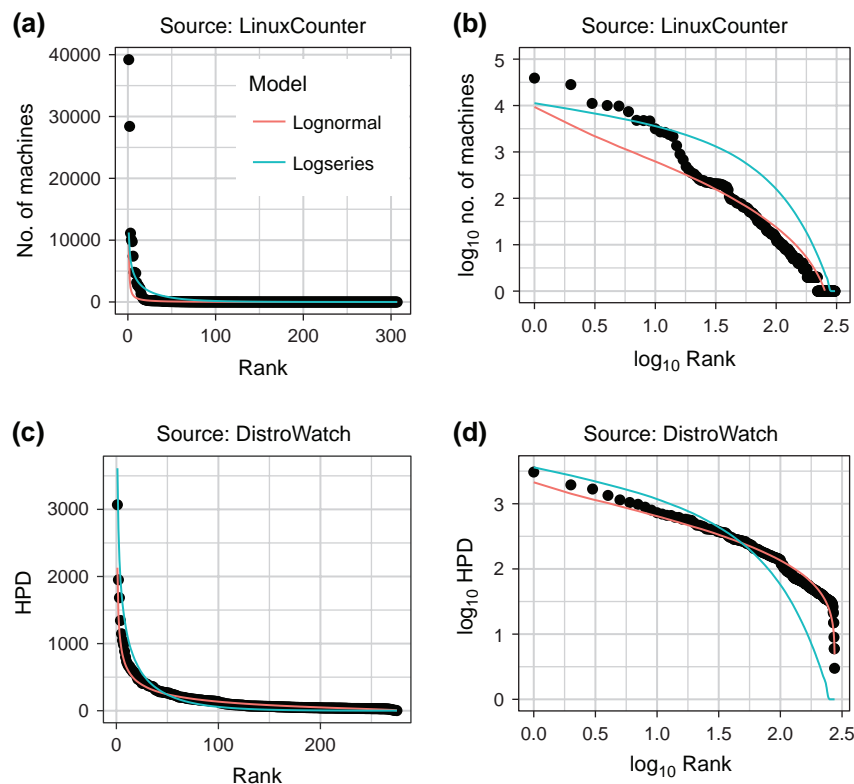


Figure 1. Rank-commonness distributions of 275 extant Linux distros from Distrowatch (a–b) and 307 distros from LinuxCounter (c–d). HPD stands for ‘hits-per-day’ on www.distrowatch.com; no. of machines is number of actual computers on which a particular distro is installed in the set of all computers registered on LinuxCounter – both are proxies for distro commonness (abundance). Black points are the data, solid lines are the lognormal (red) and logseries (blue) SAD models fitted to the data.

Calculation of Taylor's power law (TPL)

For each distro with ≥ 10 yr monitored on Distrowatch, and for each distro monitored by Wikimedia traffic reports, we calculated log-transformed temporal variance of the commonness (Distrowatch HPD or monthly Wikipedia requests) and log-transformed temporal mean of commonness. Although a non-linear regression is potentially better suited to estimate the TPL scaling exponent, we followed the practice in the broad literature on TPL and fitted a normal linear regression to the data using log variance versus log mean. We took the slope of the regression as the estimate of the TPL scaling exponent, and we hereafter call it 'TPL slope'.

Data on niche breadth

Niche and niche breadth can be complex concepts (Chase and Leibold 2003, Peterson et al. 2011), here we use a simplified definition of niche breadth as the suite of environments or resources that a species can inhabit or use (Gaston et al. 1997). We used three proxies of niche breadth of Linux distros.

1) Number of packages. We made the analogy between species functional traits and software packages that come pre-installed with each distro on an installation medium, for example as an .iso file on a live DVD. We defined the number of software packages as niche breadth – more packages mean wider niche breadth. The number of packages was extracted for each distro listed at < www.Distrowatch.com > between 28/04/2015 and 27/04/2016. Distrowatch provides detailed information on each distro, from which we extracted the full list of packages (Ubuntu-specific example: < <http://goo.gl/0Qhflk> >).

2) Number of applications. We used 32 'application' categories defining the broad purposes of each distro. These categories were Assistive, Beginners, Chromebooks, Clusters, Data Rescue, Desktop, Disk Management, Education, Firewall, Forensics, Free Software, Gaming, High Performance Computing, Live CD, Live DVD, Live Medium, Multimedia, Myth TV, Netbooks, Network Attached Storage, Old Computers, Privacy, Raspberry Pi, Router, Scientific, Security, Server, Source-Sbased, Specialist, Telephony, Thin Client, and UNIX. We use number of these applications as a measure of niche breadth. We merged Live CD, Live DVD and Live Medium to a single category. On Distrowatch, each distro can be labeled with any combination of these categories. These labels can be obtained from each distro-specific website (example of Ubuntu: < <http://goo.gl/VDMUt6> >).

3) Number of special applications. In many biodiversity datasets, patterns are mainly influenced by a small number of common species, which can hide potentially interesting patterns of the rare species (Jetz and Rahbek 2002). We applied this concept here and used the same number of applications as above, but excluded the prevalent 'Desktop' and 'Live CD' categories, so that the number reflects only the relatively narrow applications.

In total, we obtained the data for 275 distros, from which we further removed operating systems based on BSD, rolling release distros (e.g. Arch Linux), distros with more than 8000 packages, and the Debian distro. This was because package lists for these distros on Distrowatch show all available packages in their repositories, not packages that are bundled on installation media (L. Bodnár, Distrowatch administrator, pers. comm.). This left us with 227 distros for the analysis.

Relationship between niche breadth and commonness

We fit a generalized linear model (quasipoisson family, log link function) with commonness as a response and the three measures of niche breadth (described above) as predictors; we log-transformed the number of packages prior to the modeling. For commonness, we used both the HPD measure obtained from Distrowatch (227 data points) and the number of machines with a given distro obtained from Linux-Counter (108 data points).

Phylogenetic information

The majority of current GNU/Linux distros descended from three original distros: Debian, Red Hat and Soft Landing Linux system (SLS, later known as Slackware). Family trees of these three Linux families were compiled by the GNU/Linux Timeline project (GLDT) consisting of A. Lundqvist, D. Rodic, M. A. Mustafa, A. Urosevic and J. A. Sandoval; the coded trees are available at < <http://futurist.se/gldt/> > under GNU Free Documentation Licence. These trees were last updated in 2012, which is also the point at which our phylogenetic analyses terminate. We converted the family trees into Newick phylogenetic trees, and for each distro we extracted the date at which its development began ('speciation'), and the date at which development ceased ('extinction'). Here we rely on the dates provided by the GLDT project, which does not consider the possibility of renewed development after a 'dormancy' period. Although such 'de-extinctions' can happen in theory, they either did not occur, or were not coded in our three Linux phylogenies. Further, the GLDT estimates of the last development dates may be imprecise – here we assume that the imprecision is either lower, or comparable with, the imprecision of extinction dates derived from fossil data (Wang and Marshall 2016) or phylogenies (Morlon 2014), which itself can be substantial.

Speciation and extinction through time

We separated the process of diversification among GNU/Linux distros into speciation and extinction. For each of the three distro families we plotted: 1) the cumulative number of speciation events, which is the number of distros that had been created up to a given date (even if development had ceased) – analogous to cumulative speciation through time plots created for biological systems, the so-called

'lineage-through-time plots' (Nee et al. 1994). 2) The cumulative number of extinction events through time, which is the cumulative number of distros that had ceased development. 3) Per-species instantaneous speciation and extinction rates; these are numbers of distros that were created or their development ceased in a given month, divided by the total number of all extant distros in that month. 4) Per-species instantaneous diversification rate, defined as speciation rate minus extinction rate. This quantity is of key interest, since it is expected to slow down over time.

Phylogenetic signal of traits

For this analysis, we used 56 distros that descended from Debian (excluding Debian) – these are the distros for which we have both data on traits (i.e. packages) and phylogenetic relationships. We used the same data on packages as in the analysis of niche breadth (above). Data and lists for all software packages of the 56 distros were collected from < www.distrowatch.com > on 14/05/2016 (see the Material and methods section on niche breadth for more details). We created a package by distro binary matrix, with 1 for presences of the package, and 0 for absence. To test for the phylogenetic signal we performed two analyses.

1) We estimated if the dissimilarity in composition of binary traits is related to the phylogenetic distance between distros. We calculated two distance matrices of compositional dissimilarity, one using β_j (Jaccard beta), the other β_{sim} (Beta sim) (Koleff et al. 2003). We also calculated a distance matrix based on phylogenetic distance between the 56 distros. We then used Mantel tests (1000 permutations) to detect significant correlations between distance matrices representing compositional dissimilarity and phylogenetic distance. We also calculated Spearman correlations (Rho) between the matrices.

2) We measured the strength of phylogenetic signal in each trait separately. If the traits were continuous, it would have been possible to use lambda (Pagel 1999) or K (Blomberg et al. 2003) statistics. Since our traits are all binary (presence or absence of a package), we used the D statistic (Fritz and Purvis 2010) implemented in function phylo.d in R package caper. D measures phylogenetic signal in binary traits, and can have negative or positive values, with $D < 0$ indicating phylogenetically conserved traits, $D = 1$ indicating random distribution of trait states along the phylogeny, and indicating $D > 1$ overdispersed traits. We calculated D for each package, together with the probability that the observed D comes from a randomly distributed package along the phylogeny.

Data and code

All of the data and R code used for the analyses are archived at Zenodo.org < <http://doi.org/10.5281/zenodo.1120445> > under GNU General Public License, ver. 2.

Results

We found that the diversity patterns observed in the GNU/Linux universe matched different macroecological and macroevolutionary patterns to varying degrees.

Macroecological patterns

Species commonness measured as the number of machines or as HPD, followed a skewed frequency distribution (Fig. 1), with many distros being uncommon and few distros being common, which is typical for ecological abundance data. In both metrics, the distribution can be approximated by a log-normal probability density function, which performed better than the logseries model. Although the distribution of HPD lacked the typical tail of singletons (Fig. 1d), this is likely because Distrowatch does not keep track of the extremely rare distros – in this respect the data from LinuxCounter are more representative.

We found that the relationship between commonness mean (M) and variance (V) can be modelled by a power function in all three examined datasets (Fig. 2), with slope of approximately 2 (Table 2). We detected no significant positive relationship between commonness and the three measures of niche breadth (Fig. 3), with the exception of a weak but significant, positive relationship between commonness and the number of all software packages (quasipoisson deviance explained = 7%, $p < 0.01$, Fig. 3a), and between commonness and the number of all applications (quasipoisson deviance explained = 12%, $p < 0.001$, Fig. 3b) in the HPD metric from Distrowatch.

Macroevolutionary patterns

In all three Linux families (i.e. larger groups of Linux distros with the same origin), the cumulative number of phylogenetic lineages increased linearly through time (Fig. 4a, b).

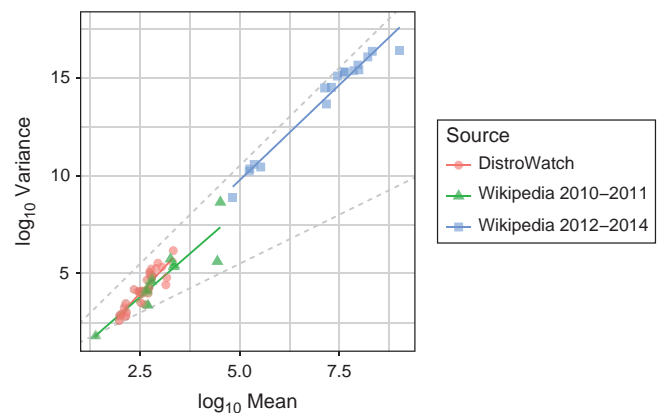


Figure 2. Taylor's power law (TPL) of temporal fluctuation of commonness of Linux distributions, estimated from three data sources. Each point represents a Linux distribution. Dashed lines have slopes 1 and 2, and are presented for visual guidance. Parameters of the linear regressions are in Table 1.

Table 2. Parameters of the Taylor's power law (TPL) describing temporal fluctuation of commonness of Linux distributions. Provided are slope and intercept of the linear regression of log temporal variance vs log temporal mean, their standard errors (SE) and the coefficient of determination (R^2).

Source	Intercept (SE)	Slope (SE)	R^2
DistroWatch	-3.34 (1.25)	2.18 (0.2)	0.8
Wikipedia 2010–2011	-1.4 (2.42)	1.7 (0.32)	0.81
Wikipedia 2012–2014	0.11 (1.43)	1.94 (0.08)	0.97

In the Debian family, there was a peak in instantaneous per-species speciation rate around 2005, followed by a pronounced slowdown and a peak of extinction rates with a subsequent slowdown in 2006 (Fig. 4c). Similar patterns occurred in the Red Hat family, but three years earlier (Fig. 4c). Diversification in all three Linux families always showed a decline after 2005 (Fig. 4d). However, in Debian and SLS there also was a relatively low rate of diversification rate prior to 2005, while Red Hat underwent a rapid diversification even prior to 2005 (Fig. 4d).

Our analyses of phylogenetic signal in trait composition showed that more closely related distros (measured by temporal distance from their nearest common ancestor) were more similar in their composition of traits measured by beta-diversity of package composition (Fig. 5b, c) than randomly selected pairs of distros. Further, 17% and 23% of the 14,161 packages exhibited phylogenetic signal significant at $\alpha = 0.05$ and $\alpha = 0.1$ respectively, measured by the D statistic (Fig. 5d, e).

Discussion

Our analyses demonstrate that the Linux universe and biological systems not only showcase many structural analogies (as already pointed out by Mens et al. 2014a, b), but also share quantitatively similar emerging properties, including patterns of commonness and evolutionary rates. This is in line with findings from other complex systems, which also exhibit macroecological patterns (Gaston et al. 1993, Mace and Pagel 1995, Bettencourt et al. 2007, Nekola and Brown 2007, Warren et al. 2011, Blonder et al. 2014); and macroevolutionary patterns (Gjesfeld et al. 2016). The Linux universe shows that not one, but a whole spectrum of biological patterns can emerge in a single non-biological system. Thus, we should expect similar patterns to emerge in other complex systems with categories of objects that evolve, for example in languages, musical styles, political parties, companies, or even countries and religions.

At this point the reader might call for a mechanistic model that could replicate, or predict, the patterns across unrelated systems. Although aimed at describing the origin of different patterns compared to this investigation, such an approach has been already adopted e.g. by Gherardi et al. (2013), where a simple stochastic model was able to predict both distribution of Linux package sizes and mammalian body masses. However, to replicate the multitude of patterns presented here, we would have to attain a degree of sophistication that goes

beyond the scope of a single paper. Instead, for now we resort to a parsimonious and non-mechanistic explanation for the prevalence of the patterns. Simply, when systems share structural constraints, such as partitioning of objects to categories (species) across space (Frank 2009, Harte 2011), or when variables in the system interact in similar ways [additively vs multiplicatively; (Blonder et al. 2014)], they will exhibit similar quantitative patterns, independent on system identity and its detailed inner working. This follows from both the central limit theorem (McGill and Nekola 2010) and the theory of maximum entropy (Harte 2011). Thus, we can explain the emergence of the patterns without any typically biological mechanism such as genetic basis for phenotypic variation, complex species interactions, behavior, or community assembly rules. In fact, all of these mechanisms are mostly absent in the Linux universe, yet the biodiversity patterns emerge anyway, which really points to their non-biological causes. In the following, we discuss each investigated pattern individually.

Macroecological patterns

We showed that commonness of GNU/Linux-based operating systems follows both a skewed frequency distribution and a power relationship between mean commonness and its temporal variance. In ecology, species-abundance distributions (Fischer 1943, McGill et al. 2007, Nekola and Brown 2007) and distributions of range sizes (Gaston 1996b, 2003) are often best described by the same skewed lognormal or logseries model (Baldrige et al. 2016). This pattern emerges when individuals within a population have roughly constant per-individual probabilities of reproducing or dying, which in turn leads to stochastic multiplicative (as opposed to additive) dynamics of population size (Lande et al. 2003). In the Linux universe, these multiplicative dynamics emerge when users install (reproduction) and uninstall distros (death) on computers, and when each user has a roughly constant probability of installing his/her favourite distro on a computer, or abandoning the distro.

Similar dynamics can also produce the mean-variance scaling of population abundance fluctuations known as Taylor's power law (TPL), which we also detected in the Linux universe. TPL generally has a proportion of explained variance of > 0.8 in both single- and multi-species systems (Taylor and Woivod 1980, 1982, Hubbell 2001), with the slope typically falling between 1 and 2 (Kendal 2004, Keil et al. 2010). Multiple hypotheses have been suggested to interpret empirical TPL slopes (Kendal 2004). Slopes closer to 1 were linked to reproductive asynchrony (Ballantyne and Kerkhoff 2007), species interactions (Kilpatrick and Ives 2003), hard upper limits on population size (Keil et al. 2010), and even to statistical artifacts (Kiflawi et al. 2016), while slopes closer to 2 emerge from simple multiplicative models such as random walk or deterministic chaos (Perry 1994). The slopes close to 2 observed here suggests that the temporal dynamics of Linux commonness follow an unbounded stochastic multiplicative model, which is also what we expect to lead to the observed skewed frequency distributions.

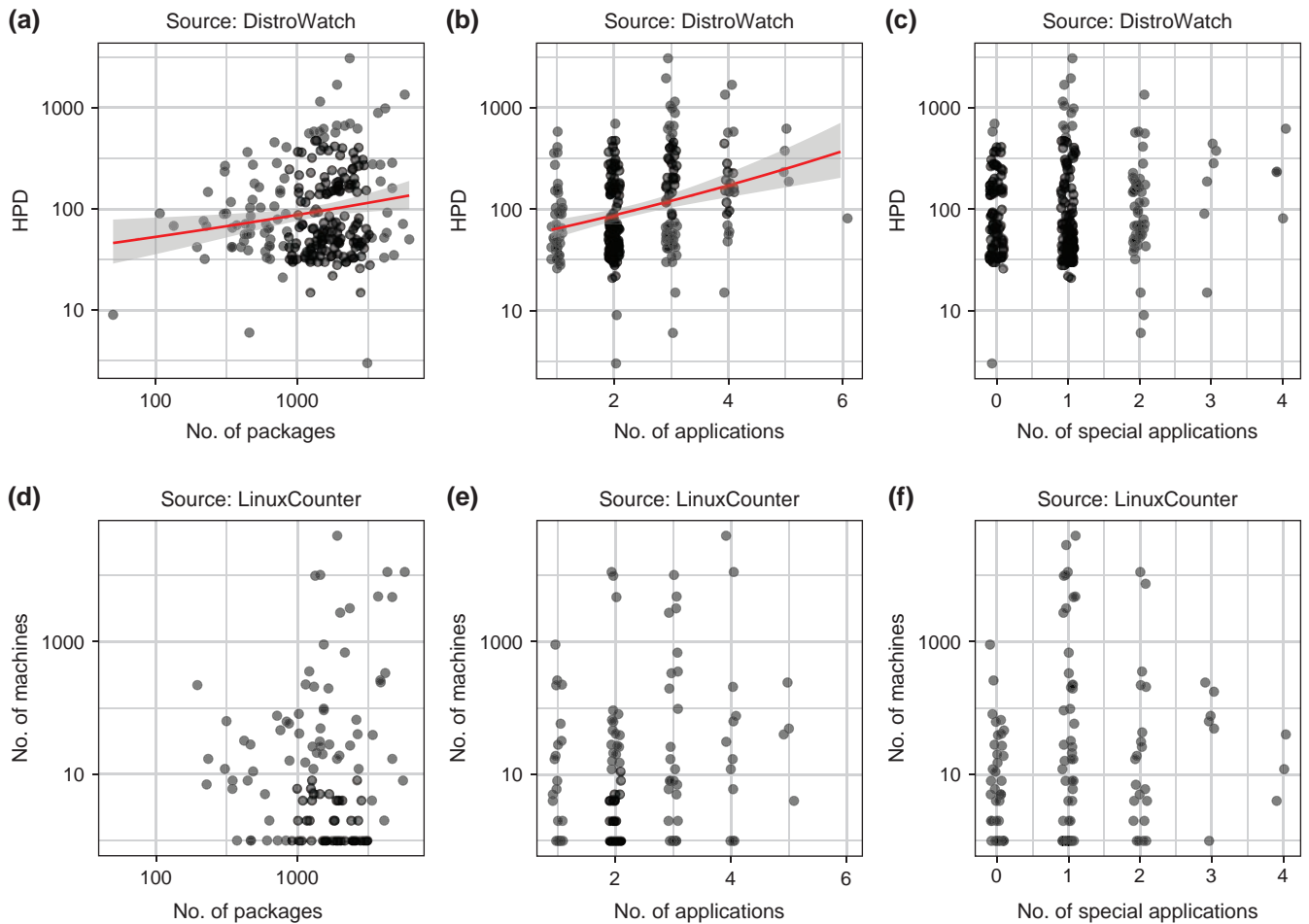


Figure 3. Three measures of niche breadth and their relationship with commonness measured as HPD or no. of machines. (a, d) Niche breadth measured as number of packages, (b, e) niche breadth measured as number of all applications, (c, f) niche breadth measured as number of special applications (not including ‘Desktop’ and ‘Live CD’). Solid red lines are means of quasipoisson log-linear regressions, shading shows standard errors of the means. HPD stands for ‘hits-per-day’ on <www.distrowatch.cz>; no. of machines is the number of actual machines on which a particular distro is installed in the set of machines registered on LinuxCounter – both are proxies for distro commonness (abundance).

We detected no, or only very weak, positive relationships between commonness and niche breadth of distros (i.e. the functionality that distros offers), both when measuring it as the number of software packages in a distro or as the number of applications (as stated by the developers). In ecology, such positive correlation between niche breadth and commonness (specifically, geographical range size) is a general pattern (Slatyer et al. 2013). The hypothesized ecological explanation is that by exploiting a greater array of resources and maintaining populations in a wider variety of conditions, species may become more common (Brown 1984). However, our findings fail to generalize this ecological relationship to the Linux universe. A possible reason for the absence (or weakness) of the link could be our definition of niche breadth, as the number of software packages that come pre-installed with a distro can be easily expanded. Unlike biological phenotypes, which are usually fixed over ecological timescales, numerous additional packages are available for many distros that can

be custom-added at any time, and this likely weakens the strength of using pre-installed software for determining the success of an operating system.

Macroevolutionary patterns

Diversification slowdowns are often observed in biological systems [Rabosky and Glor (2010), Moen and Morlon (2014), but see Harmon and Harrison (2015) and Graham et al. (2016)], and can be explained by competition for limited resources or space. Similar reasoning can also explain the diversification slowdowns that we detected in the Linux universe: we suggest that between years 2000 and 2005 Linux users had become comfortable with certain distros that satisfied most of the potential applications and user requirements and therefore diversification slowed down. An alternative explanation can be borrowed from (Gjesfeld et al. 2016) who observed diversification slowdowns in American

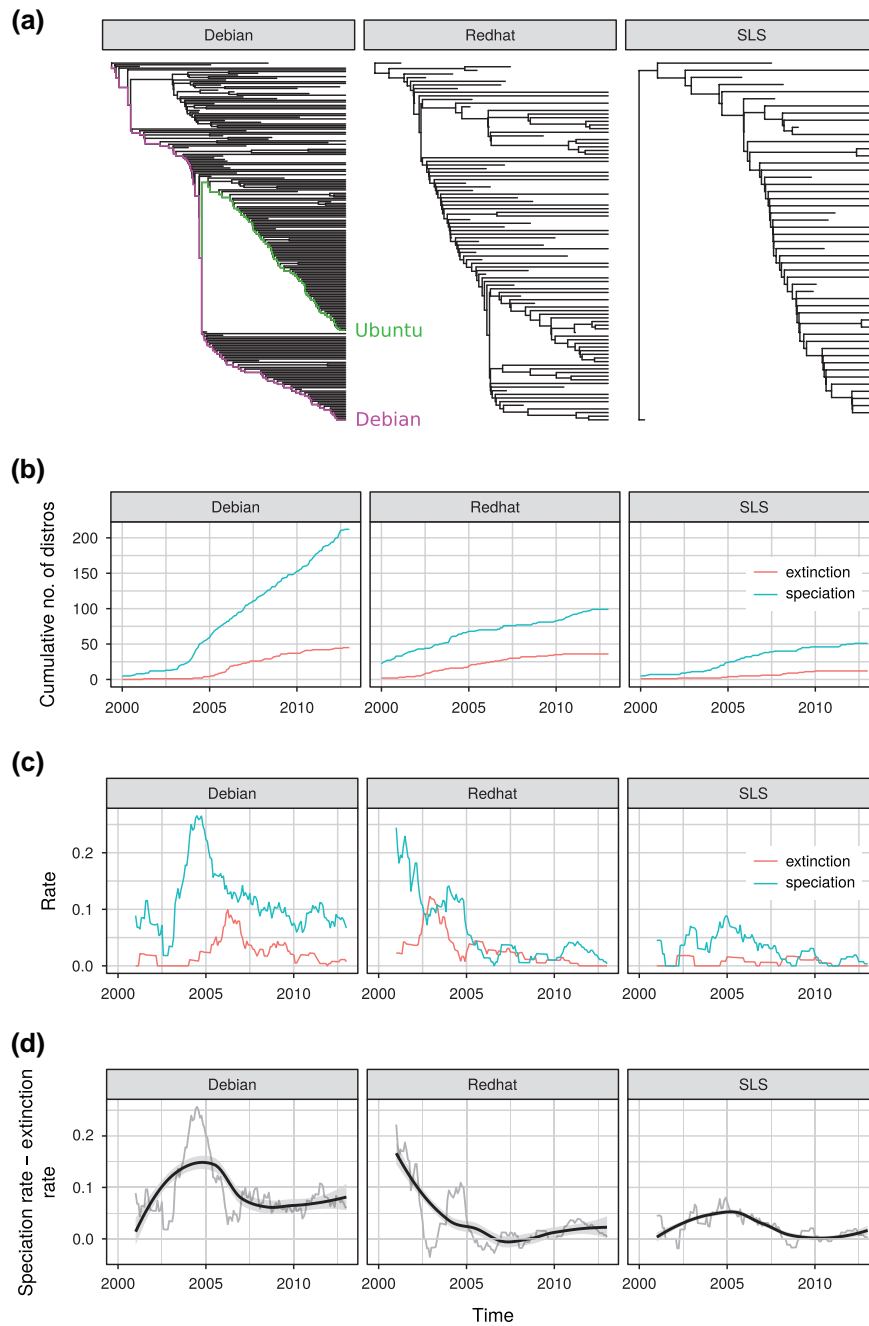


Figure 4. Speciation, extinction, and diversification of Linux distros through time in three major distro families. (a) Phylogenetic trees of the three families of Linux distros that were used to make the panels (b), (c), and (d). Branches of two major distros, Ubuntu and Debian, are highlighted by green and purple. (b) Accumulation of new species (distros) and extinctions over time. (c) Instantaneous per-distro rates of speciation and extinction rate. (d) Instantaneous per-distro diversification rate (speciation rate – extinction rate). Solid black lines are LOWESS regressions (smoothing span 0.3), shading indicates standard errors of the smoothed means.

automobiles, and who argue that the slowdowns can be explained by the emergence of dominant designs (Abernathy and Utterback 1978). As Gjesfeld et al. (2016) explains, when technologies become too specialized, the cost of implementing innovations becomes too high and diversification slows down, leading to long-term dominance of the most successful designs. However, since the community-driven

universe of Linux is open and largely free from the usual cost or copyright constraints, we consider this economical explanation implausible, and we tend to lean towards explaining the slowdowns by the ecological mechanism involving saturation of a finite niche space.

As mentioned, Linux distros come with diverse functionalities, which are pre-installed in the form of software

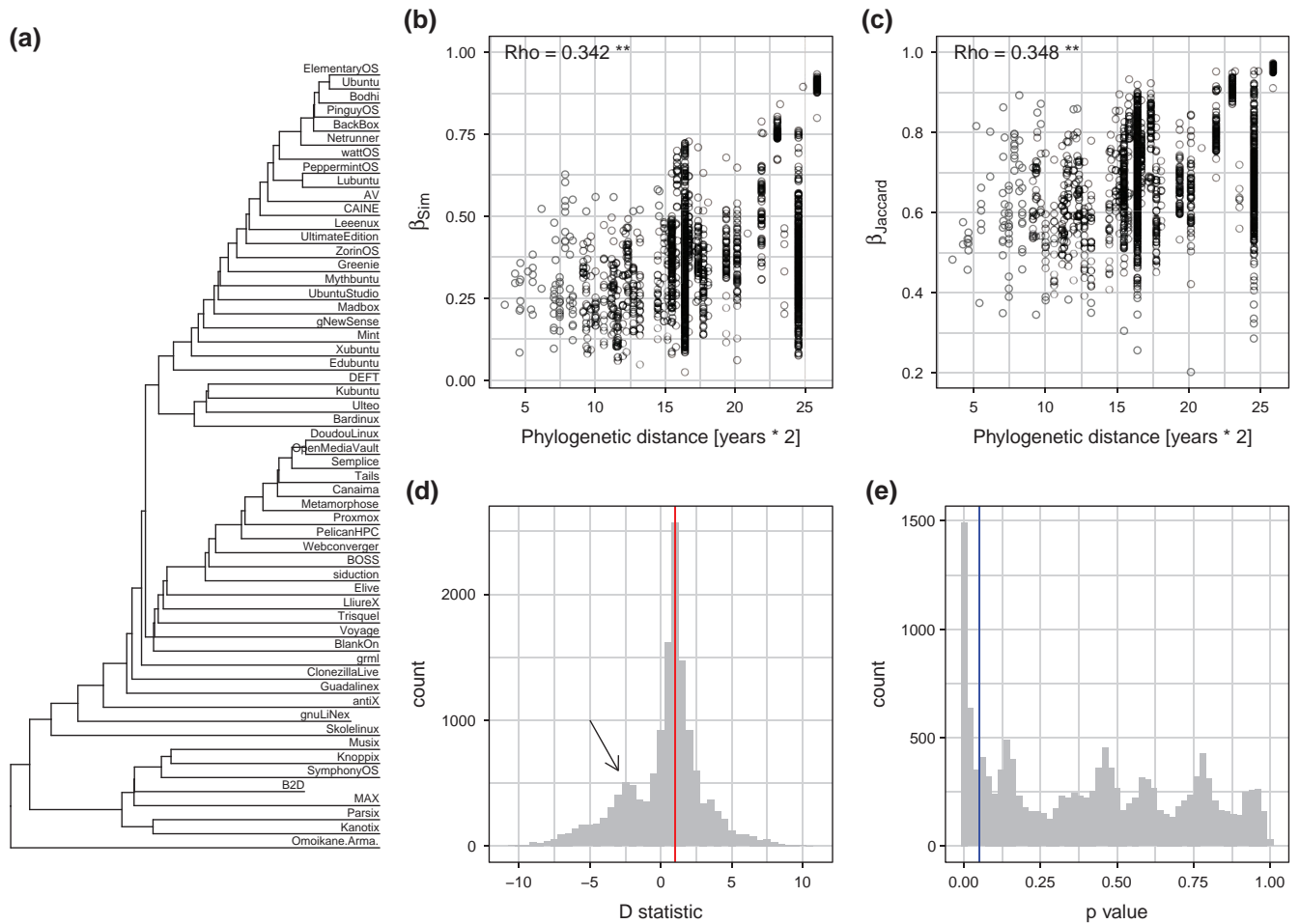


Figure 5. Phylogenetic signal in 14 161 binary traits (software packages) of 56 Debian-based distros for which we had information on package composition. (a) Phylogenetic tree of the distros. (b–c) Phylogenetic signal in the composition of software packages shipped with each distro, expressed as (significant) positive correlation between phylogenetic distance and trait compositional dissimilarity measured by $\beta_{Jaccard}$ and β_{sim} . Rho is Spearman correlation, asterisks indicate probability $p < 0.01$ that there is no correlation (Mantel test, 9999 permutations). (d) Frequency distribution of the D statistic (Fritz and Purvis 2010) across the 14 161 binary traits; red line ($D = 1$) indicates random distribution of a trait along the phylogeny, $D < 0$ is phylogenetic conservatism. (e) Distribution of probability that a given D comes from a trait randomly distributed along the phylogeny; blue line is $p = 0.05$.

packages, and which we have made analogous to functional traits of biological species. We found that more closely related distros were more similar in their composition of traits than randomly selected pairs of distros, and 17 to 23% of individual software packages were significantly phylogenetically conserved. These results are consistent with a widespread macroevolutionary pattern known as phylogenetic ‘signal’ or ‘autocorrelation’ of traits, i.e. the tendency of more closely related species to be ecologically (functionally) more similar than species drawn at random from the phylogenetic tree of closely related species (Harvey and Pagel 1991, Wiens et al. 2010).

Finding a phylogenetic signal in Linux distros is expected. Some degree of functional similarity of a descendant distro with the parent is indeed desired, perhaps by the users (they like the familiar), by the developers (they like to build on what has already been built), as well as by the need to

avoid radical changes that cause bugs due to complex package dependencies (Mens et al. 2014a, b). At the same time, the signal is not particularly strong. The likely reason for the weakening is that distros are not fully confined to directly inherit packages from their ancestors. Instead, developers of distros are free to load them with existing open-source software that was originally developed for other distros, making the process of package inheritance more similar to horizontal gene transfer in bacteria (Ochman et al. 2000) or to horizontal transmission of culture in human societies (Henrich and Broesch 2011, Hewlett et al. 2011).

Linux as a model system for biodiversity

Being well-documented, Linux may serve as a useful model system for studying certain macroecological or macroevolutionary patterns and dynamics that are particularly hard to

get by, given our limited capacity to document biological processes across space, time, and taxa (Hortal et al. 2015, Meyer et al. 2015), and the difficulty of conducting experiments over large spatial and temporal scales. For instance, we unveiled slowdowns in both diversification and extinction rates in Linux, processes that happen at timescales that are impossible to directly observe in nature, and usually need to be indirectly inferred using strong assumptions and complex models.

The potential use of the Linux universe as a model for biodiversity may indeed extend beyond the comparisons made here. Here, we have made the first steps, and these initial analogies can be complemented with a more thorough exploration of the Linux system to fully assess their utility for macroecological and macroevolutionary research. More analogies could be made, and patterns analyzed, based on easily retrievable empirical data on Linux. For instance, economic activity tied to Linux applications could serve as an analogy to productivity of geographical areas and allow studying the similarity in productivity–biodiversity relationships (Currie et al. 2004) of the two systems. Another example is the positive relationship between area and speciation rate (Lomolino et al. 2010, Wagner et al. 2014): populations in larger areas are more likely to become isolated and drift apart genetically, or to encounter a larger variety of selection pressures and undergo adaptive radiations. Our preliminary analyses suggest that such a relationship can be observed in the Linux universe (Supplementary material Appendix 1 Fig. A1).

Towards cultural ecology

Darwinian evolutionary thinking and models have successfully been adapted to study evolution of cultural systems (Dawkins 1976, Cavalli-Sforza and Feldman 1981, Boyd and Richerson 1985, Sperber 1996, Mesoudi 2016, 2017). However, as shown in our results and elsewhere, there is a broader variation of culture that also has analogies in ecology, and that exhibits ecological patterns. Here we see potential for a new discipline of cultural ecology, which would use ecological models to explain culture. For example, fluctuation and distribution of rarity and commonness can be used to predict extinction risk (Kunin and Gaston 1997) in cultures and technologies. There is also a potential for biodiversity theory, and its popular models, such as the neutral model (Hubbell 2001) or the theory of island biogeography (MacArthur and Wilson 1967), to predict and map patterns of cultural diversity. We suggest that one aspect of cultural variation that can benefit from biodiversity theory, is the issue of cultural and technological diversity (e.g. in nations, towns, or companies) and its relationship with productivity and stability – a large body of relevant theory can be offered here by the biodiversity–ecosystem function research (BEF; Loreau et al. 2002). Finally, some ecological models, such as the neutral theory of biodiversity (Hubbell 2001), offer direct connections to evolutionary theory, and thus to an even broader interdisciplinary integration.

Acknowledgements – We are grateful to Luke Harmon, Antonín Macháč and eight anonymous referees for extensive comments. Ladislav Bodnář from <www.distrowatch.com> assisted with access to the raw data.

Funding – This study was conceived at a workshop (25–29/4/2016, Aix-en-Provence, France) organized by Alison Specht and the French Center for Synthesis and Analysis of Biodiversity (CESAB), together with Marten Winter and the synthesis division (sDiv) of the German Centre for Integrative Biodiversity Research (iDiv), funded via the German Research Foundation (DFG FZT 118). BY benefited from a postdoctoral fund from the French national research agency LabEx ANR-10-LABX-0003-BCDiv, in the context of the ‘Investissements d’avenir’ no. ANR-11-IDEX-0004-02. KSR was supported by ERC Adv grant 26055290. CM acknowledges support from the Volkswagen Foundation through a Freigeist Fellowship.

References

- Abernathy, W. J. and Utterback, J. M. 1978. Patterns of industrial innovation. – *Technol. Rev.* 80: 40–47.
- Baldrige, E. et al. 2016. An extensive comparison of species-abundance distribution models. – *PeerJ* 4: e2823.
- Ballantyne, F. and Kerkhoff, A. J. 2007. The observed range for temporal mean-variance scaling exponents can be explained by reproductive correlation. – *Oikos* 116: 174–180.
- Bettencourt, L. M. A. et al. 2007. Growth, innovation, scaling, and the pace of life in cities. – *Proc. Natl Acad. Sci. USA* 104: 7301–7306.
- Blomberg, S. P. et al. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. – *Evolution* 57: 717–745.
- Blonder, B. et al. 2014. Separating macroecological pattern and process: comparing ecological, economic, and geological systems. – *PLoS One* 9: e112850.
- Bonds, M. H. et al. 2012. Disease ecology, biodiversity, and the latitudinal gradient in income. – *PLoS Biol.* 10: e1001456.
- Boyd, R. and Richerson, P. J. 1985. *Culture and the evolutionary process*. – Univ. of Chicago Press.
- Brown, J. H. 1984. On the relationship between abundance and distribution of species. – *Am. Nat.* 124: 255–279.
- Cavalli-Sforza, L. L. and Feldman, M. W. 1981. *Cultural transmission and evolution: a quantitative approach*. – Princeton Univ. Press.
- Chase, J. M. and Leibold, M. A. 2003. *Ecological niches: linking classical and contemporary approaches*. – Univ. of Chicago Press.
- Currie, D. J. et al. 2004. Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. – *Ecol. Lett.* 7: 1121–1134.
- Darwin, C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. – John Murray.
- Dawkins, R. 1976. *The selfish gene*. – Oxford Univ. Press.
- Drakare, S. et al. 2006. The imprint of the geographical, evolutionary and ecological context on species–area relationships. – *Ecol. Lett.* 9: 215–227.
- Edge, J. 2011. Distribution “popularity.” – <<https://lwn.net/Articles/470782/lwn.net>>.
- Fischer, R. A. 1943. A theoretical distribution for the apparent abundance of different species. – *J. Anim. Ecol.* 12: 54–58.

- Fortuna, M. A. et al. 2011. Evolution of a modular software network. – *Proc. Natl Acad. Sci. USA* 108: 19985–19989.
- Frank, S. A. 2009. The common patterns of nature. – *J. Evol. Biol.* 22: 1563–1585.
- Fritz, S. A. and Purvis, A. 2010. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. – *Conserv. Biol.* 24: 1042–1051.
- Gaston, K. J. 1996a. The multiple forms of the interspecific abundance–distribution relationship. – *Oikos* 76: 211–220.
- Gaston, K. J. 1996b. Species-range-size distributions: patterns, mechanisms and implications. – *Trends Ecol. Evol.* 11: 197–201.
- Gaston, K. J. 2000. Global patterns in biodiversity. – *Nature* 405: 220–227.
- Gaston, K. J. 2003. The structure and dynamics of geographic ranges. – Oxford Univ. Press.
- Gaston, K. J. et al. 1993. Comparing animals and automobiles: a vehicle for understanding body size and abundance relationships in species assemblages? – *Oikos* 66: 172–179.
- Gaston, K. J. et al. 1997. Interspecific abundance–range size relationships: an appraisal of mechanisms. – *J. Anim. Ecol.* 66: 579–601.
- Gherardi, M. et al. 2013. Evidence for soft bounds in Ubuntu package sizes and mammalian body masses. – *Proc. Natl Acad. Sci. USA* 110: 21054–21058.
- Gjesfeld, E. et al. 2016. Competition and extinction explain the evolution of diversity in American automobiles. – *Palgrave Commun.* 2: 16019.
- Graham, C. H. et al. 2016. Phylogenetic scale in ecology and evolution. – *bioRxiv*: 063560.
- Harmon, L. J. and Harrison, S. 2015. Species diversity is dynamic and unbounded at local and continental scales. – *Am. Nat.* 185: 584–593.
- Harte, J. 2011. Maximum entropy and ecology: a theory of abundance, distribution, and energetics. – Oxford Univ. Press.
- Harvey, P. H. and Pagel, M. D. 1991. The comparative method in evolutionary biology. – Oxford Univ. Press.
- Henrich, J. and Broesch, J. 2011. On the nature of cultural transmission networks: evidence from Fijian villages for adaptive learning biases. – *Phil. Trans. R. Soc. B* 366: 1139–1148.
- Hewlett, B. S. et al. 2011. Social learning among Congo Basin hunter-gatherers. – *Phil. Trans. R. Soc. B* 366: 1168–1178.
- Hortal, J. et al. 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. – *Annu. Rev. Ecol. Evol. Syst.* 46: 523–549.
- Hubbell, S. P. 2001. The unified theory of biodiversity and biogeography. – Princeton Univ. Press.
- Jetz, W. and Rahbek, C. 2002. Geographic range size and determinants of avian species richness. – *Science* 297: 1548–1551.
- Keil, P. et al. 2010. Predictions of Taylor's power law, density dependence and pink noise from a neutrally modeled time series. – *J. Theor. Biol.* 265: 78–86.
- Kendal, W. S. 2004. Taylor's ecological power law as a consequence of scale invariant exponential dispersion models. – *Ecol. Complex.* 1: 193–209.
- Kiflawi, M. et al. 2016. Heterogeneous 'proportionality constants' – a challenge to Taylor's power law for temporal fluctuations in abundance. – *J. Theor. Biol.* 407: 155–160.
- Kilpatrick, A. M. and Ives, A. R. 2003. Species interactions can explain Taylor's power law for ecological time series. – *Nature* 422: 65–68.
- Koleff, P. et al. 2003. Measuring beta diversity for presence–absence data. – *J. Anim. Ecol.* 72: 367–382.
- Kunin, W. E. and Gaston, K. J. 1997. The biology of rarity: causes and consequences of rare common differences. – Chapman and Hall.
- Lande, R. et al. 2003. Stochastic population dynamics in ecology and conservation. – Oxford Univ. Press.
- Lawton, J. H. 1999. Are there general laws in ecology? – *Oikos* 84: 177–192.
- Lewontin, R. C. 1970. The units of selection. – *Annu. Rev. Ecol. Syst.* 1: 1–18.
- Lomolino, M. V. et al. 2010. Biogeography. – Sinauer Associates.
- Loreau, M. et al. 2002. Biodiversity and ecosystem functioning. – Oxford Univ. Press.
- Losos, J. B. 2008. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. – *Ecol. Lett.* 11: 995–1003.
- MacArthur, R. H. and Wilson, E. O. 1967. The theory of island biogeography. – Princeton Univ. Press.
- Mace, R. and Pagel, M. 1995. A latitudinal gradient in the density of human languages in North America. – *Proc. R. Soc. B* 261: 117–121.
- Malthus, T. R. 1798. An essay on the principle of population. – J. Johnson.
- Maynard Smith, J. 1982. Evolution and the theory of games. – Cambridge Univ. Press.
- McGill, B. J. and Nekola, J. C. 2010. Mechanisms in macroecology: AWOL or purloined letter? Towards a pragmatic view of mechanism. – *Oikos* 119: 591–603.
- McGill, B. J. et al. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. – *Ecol. Lett.* 10: 995–1015.
- Mens, T. et al. 2014a. Studying evolving software ecosystems based on ecological models. – In: Mens, T. et al. (eds), *Evolving software systems*. Springer, pp. 297–326.
- Mens, T. et al. 2014b. ECOS: ecological studies of open source software ecosystems. – *Software Maintenance, Reengineering and Reverse Engineering (CSMR-WCRE)*, pp. 403–406.
- Mesoudi, A. 2016. Cultural evolution: a review of theory, findings and controversies. – *Evol. Biol.* 43: 481–497.
- Mesoudi, A. 2017. Pursuing Darwin's curious parallel: prospects for a science of cultural evolution. – *Proc. Natl Acad. Sci. USA* 114: 7853–7860.
- Meyer, C. et al. 2015. Global priorities for an effective information basis of biodiversity distributions. – *Nat. Commun.* 6: 8221.
- Moen, D. and Morlon, H. 2014. Why does diversification slow down? – *Trends Ecol. Evol.* 29: 190–197.
- Morlon, H. 2014. Phylogenetic approaches for studying diversification. – *Ecol. Lett.* 17: 508–525.
- Morlon, H. et al. 2009. Taking species abundance distributions beyond individuals. – *Ecol. Lett.* 12: 488–501.
- Myers, C. R. 2003. Software systems as complex networks: structure, function, and evolvability of software collaboration graphs. – *Phys. Rev. E* 68: 046116.
- Nee, S. et al. 1994. The reconstructed evolutionary process. – *Phil. Trans. R. Soc. B* 344: 305–311.
- Nekola, J. C. and Brown, J. H. 2007. The wealth of species: ecological communities, complex systems and the legacy of Frank Preston. – *Ecol. Lett.* 10: 188–196.
- Ochman, H. et al. 2000. Lateral gene transfer and the nature of bacterial innovation. – *Nature* 405: 299–304.

- Pagel, M. 1999. Inferring the historical patterns of biological evolution. – *Nature* 401: 877–884.
- Pang, T. Y. and Maslov, S. 2013. Universal distribution of component frequencies in biological and technological systems. – *Proc. Natl Acad. Sci. USA* 110: 6235–6239.
- Perry, J. N. 1994. Chaotic dynamics can generate Taylor's power law. – *Proc. R. Soc. B* 257: 221–226.
- Peterson, A. T. et al. 2011. Ecological niches and geographic distributions. – Princeton Univ. Press.
- Rabosky, D. L. and Lovette, I. J. 2008 Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? – *Evolution* 62: 1866–1875.
- Rabosky, D. L. and Glor, R. E. 2010. Equilibrium speciation dynamics in a model adaptive radiation of island lizards. – *Proc. Natl Acad. Sci. USA* 107: 22178–22183.
- Rosenzweig, M. L. 1995. Species diversity in space and time. – Cambridge Univ. Press.
- Scheffer, M. et al. 2017. Inequality in nature and society. – *Proc. Natl Acad. Sci. USA* 114: 13154–13157.
- Sizling, A. L. et al. 2009. Species abundance distribution results from a spatial analogy of central limit theorem. – *Proc. Natl Acad. Sci. USA* 106: 6691–6695.
- Slatyer, R. A. et al. 2013. Niche breadth predicts geographical range size: a general ecological pattern. – *Ecol. Lett.* 16: 1104–1114.
- Sperber, D. 1996. Explaining culture: a naturalistic approach. – Blackwell.
- Taylor, L. R. and Woiwod, I. P. 1980. Temporal stability as a density-dependent species characteristic. – *J. Anim. Ecol.* 49: 209–224.
- Taylor, L. R. and Woiwod, I. P. 1982. Comparative synoptic dynamics. I. Relationships between inter- and intra-specific spatial and temporal variance/mean population parameters. – *J. Anim. Ecol.* 51: 879–906.
- Valverde, S. and Sole, R. V. 2015. Punctuated equilibrium in the large-scale evolution of programming languages. – *J. R. Soc. Interface* 12: 20150249.
- Wagner, C. E. et al. 2014. Cichlid species–area relationships are shaped by adaptive radiations that scale with area. – *Ecol. Lett.* 17: 583–592.
- Wang, S. C. and Marshall, C. R. 2016. Estimating times of extinction in the fossil record. – *Biol. Lett.* 12: 20150989.
- Warren, R. J. et al. 2011. Universal ecological patterns in college basketball communities. – *PLoS One* 6: e17342.
- Wiens, J. J. et al. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. – *Ecol. Lett.* 13: 1310–1324.
- Worster, D. 1994. Nature's economy: a history of ecological ideas. – Cambridge Univ. Press.
- Yan, K.-K. et al. 2010. Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. – *Proc. Natl Acad. Sci.* 107: 9186–9191.

Supplementary material (Appendix ECOG-03424 at < www.ecography.org/appendix/ecog-03424 >). Appendix 1.