Information Theory Broadens the Spectrum of Molecular Ecology and Evolution

Information Theory Broadens the Spectrum of Molecular Ecology and Evolution

W.B. Sherwin [1 *]

A. Chao [2]

L. Jost [3]

P.E. Smouse [4]

Addresses:

[1] Evolution and Ecology Research Centre, School of Biological Earth and Environmental Science, University of New South Wales, Sydney NSW 2052 AUSTRALIA.

and

Murdoch University Cetacean Research Unit, Murdoch University, South Road, Murdoch, WA 6150 AUSTRALIA.

[2] Institute of Statistics, National Tsing Hua University, Hsin-Chu 30043, Taiwan.

[3] EcoMinga Foundation, Via a Runtun, Baños, Tungurahua, Ecuador

[4] Department of Ecology, Evolution and Natural Resources, School of Environmental and Biological Sciences, Rutgers University, New Brunswick, NJ 08901-8551, USA,

* Correspondence W.Sherwin@unsw.edu.au

Keywords: Shannon; diversity-profile; entropy; selection; linkage-disequilibrium; gene-expression

Word-length for main text (excluding comments such as [insert fig ** here]): 3422

Highlights

Diversity of molecules or species is best expressed in terms of a diversity profile

Such profiles are useful in studies spanning bioinformatics to physical landscapes

Shannon information is a neglected but particularly informative part of the profile

Shannon now has robust theoretical background for molecular ecology and evolution

Abstract

Information or entropy analysis of diversity is used extensively in community ecology, and has recently been exploited for prediction and analysis in molecular ecology and evolution. Information measures belong to a spectrum (or '*q*-profile') of measures whose contrasting properties provide a rich summary of diversity, including allelic richness (*q*=0), Shannon information (*q*=1), and heterozygosity (*q*=2). We present the merits of information measures for describing and forecasting molecular variation within and among groups, comparing forecasts with data, and evaluating underlying processes such as dispersal. Importantly, information measures directly link causal processes and divergence outcomes, have straightforward relationship to allele frequency differences (including monotonicity that *q*=2 lacks), and show additivity across hierarchical layers such as ecology, behaviour, cellular processes, and non-genetic inheritance.

An Information Theory Base for Evolutionary Genetics and Ecology

[Insert glossary sidebar about here]

Evolutionary ecology aims to make and test forecasts about the behaviour of variants, at all levels from molecular through species to landscapes, but until recently this field has paid little attention to one of three main ways for doing this. Shannon **information (or entropy) theory** (see Glossary) is widely-used to predict macro-patterns from micro-parts, by methods such as finding the model that maximises the entropy [1], in fields as diverse as community ecology [2] and energy generation, where these methods compare favourably with alternative predictions [3]. Genes obviously carry information – about evolutionary history, recent demography, and possible future

trajectories [4, 5] – but information theory has rarely been used to investigate molecular evolution [6-9]. Shannon's information index ${}^1H$ (=H) [10], a fundamental component of information theory, is the most commonly used abundance-sensitive measure of species diversity within a community [11]:

$$ {}^1H = -\sum_{i=1}^{S} p_i \ln p_i \qquad \text{Equation 1} $$

where p_i is the proportional abundance of the i^{th} species in a community of S different species (Tutorial Box 1, and Supplement Box S1, whose footnote has definitions of all symbols). ${}^1H$ also applies to a population containing genetic variants (such as allelic types) [12], and can be thought of as the ability to spell out different messages by rearranging individual alleles (Box 1-I) – with higher diversity, a greater range of messages can be spelt out. More formally, higher ${}^1H$ means that there is reduced certainty about what type to expect when a single allele is randomly sampled.

[Insert Box 1 about here]

Historically, molecular ecology has quantified diversity with two alternative entropies (Boxes 1,S1): ${}^0H = S-1$; and ${}^2H$ (Boxes 1,S1). ${}^2H$ is the chance of choosing two different allelic types from the population, called 'heterozygosity' (${}^2H = H_e$, or for species in communities, called Gini-Simpson and variants) [4, 5, 13].

$$ {}^2H = 1 - \sum_{i=1}^{S} p_i^2 \qquad \text{Equation 2} $$

With higher diversity, the chance of randomly choosing two different alleles increases, so ${}^2H$ increases.

To create the 'q-profile' or 'spectrum', one converts (${}^{q}H = {}^{0}H, {}^{1}H, {}^{2}H,$ etc.) to a common scale of 'effective numbers' (${}^{q}D = {}^{0}D, {}^{1}D, {}^{2}D$ etc), which represent the number of equally-frequent alleles that would be needed to yield the observed ${}^{q}H$ in the sample, which typically contains alleles at unequal frequencies [14-16] (Boxes 1-II,S1,S2). The ${}^{q}D$-measures are sometimes called 'true diversity', and the use of the ${}^{q}D$ profile has long been recommended in ecology, because each q-value provides different insights into the composition of the diversity [17, 18]. For example, higher q-values emphasise the more common variants (Box 1-I,II). As well as these 'alpha'-measures for alleles in a single population, each q-value has measures of diversification among locations, which are the beta-measures (Box 1-III) that are critical in evolution and conservation [19, 20]. Finally, the total diversity within and among localities is labelled gamma diversity.

Partial ${}^{q}D$ profiles are already used in evolution and ecology, to assess community response to changed conditions [19], as well as possible correlations between species- and gene-diversity [21]. Also, methods to infer selection or population-size changes exploit the way these processes have different effects on number of variants S (q=0) and heterozygosity (q=2) [22-24]). Many recent publications include (q=1) (Box S3), yet few authors [25-28] have exploited systematic q-profiles (Box 1-II), to obtain the power described further below.

The diversity profile is most useful if we can forecast its shape under specified histories of selection, population size, dispersal, etc. We can then either test departures from those predictions and/or

estimate forces that underlie the diversity patterns we encounter. Classically, (q=2) theory predicts values of within- and among-population measures, including heterozygosity ${}^2H$, *Jost-D*, and F_{ST}, under conditions such as neutrality, selection, subdivision, dispersal, altered population size [4, 5, 29]. Below we show that after a slow start [30-34], (q=1) predictive theory is now catching up to (q=2) theory [12, 35-37], and ${}^1H$ has been proposed as a primary measure of evolvability [38, 39].

As a reading guide, the initial sections outline how (q=1) molecular information can now predict and measure processes such as adaptation and dispersal (with additional detail in supplements). Those wanting less technical detail might skip directly to the penultimate section evaluating strengths and weaknesses of each element of the q-profile , such as the effect of q on sensitivity to rare alleles (Box 1-I), and the poor performance of some conventional (q=2) measures for analysis of among-population (β) differentiation.

We consider both neutral and adaptive genetic variants, and also discuss haploid, diploid and clonal organisms. We consider genes ('loci') with only two variants ('alleles'), such as typical SNPs (single-nucleotide polymorphisms), as well as multiallelic loci, evolving under either the SMM (stepwise mutation model - some microsatellite loci) or the IAM (infinite alleles model for 'haplotypes' or 'haptigs' [40] of variants linked on the same DNA molecule [5]). We also discuss continuous traits determined by variation of multiple genes.

Measuring and Predicting Shannon's Information for Neutral Alleles

Within-population (alpha) equations have recently been derived to predict Shannon entropy ${}^1H$ and diversity ${}^1D$, within-populations (alpha), for neutral genetic information (having no effect on adaptation), for equilibrium with constant effective population size N_e and mutation rate μ (Box S1). The new equations show good fit to both simulated and real data sets for loci evolving under several mutation processes - SNP [36], SMM [12, 37] and IAM [12, 37]. Boxes S1 and S6 show how to calculate ${}^1H$, to compare data with theoretical predictions. For example, in a rainforest tree *Elaeocarpus sedentarius*, calculated and predicted ${}^1H$ agreed [12, 41]. There are many recent examples of the empirical and theoretical use of Shannon measures to assess how molecular diversity is affected by factors such as: endangerment and bottlenecks; subdivision; environmental gradients; pedigree inbreeding; and invasions expansions introductions and host jumps (Box S3). In most cases, measures with different *q*-values yield similar interpretations, but we later discuss informative cases where they differ.

Beta differentiation of molecular diversity among populations, species, and landscapes can be summarised using measures of each order of *q*. The (*q*=0) β-measures are based on the proportion of allelic types that that are not shared by two populations (Box S1). *Jost-D* is a *q*=2 β-measure, and there are related measures such as F_{ST} [29, 42] and pairwise comparison algorithms, including STRUCTURE [43] and AMOVA [44] (Box S1). The (*q*=1) measures were explicitly designed to be hierarchical [1, 10], so partitioning of molecular diversity is particularly easy, leading to differentiation measures Mutual Information I, and Shannon Differentiation. These can be derived from a contingency table of differentiation of allele frequency between geographic locations (Boxes 2,S1) [1, 12, 33, 35, 37]. Box 2 shows how dimensions can be added to

142 incorporate diversity within and among different habitats, landscapes, etc. [35, 45-47], because log-
143 linear χ^2 (or I) is completely additive [48].

144 [Insert Box 2 about here]

145 For any pair of populations, lower dispersal, smaller population size, or greater elapsed time since
146 separation will increase molecular differentiation (and hence, mutual information I). At
147 equilibrium, simulation results have been used to derive an inverse relationship between mutual
148 information (I, q=1) and effective dispersal rate ($N_e m$), over a very wide range of effective
149 population sizes ($N_e \geq 10$) and dispersal rates ($0.001 \leq m \leq 0.30$ *i.e.* 30% dispersing per
150 generation) [12] (Box S1). This equation can be used to estimate dispersal from genetic data, and
151 outperforms predictions based on (q=2) in simulation studies (Figure 1A), as well as in laboratory
152 colonies of *Drosophila* with known $N_e m$ [12]. The waratah *Telopea speciosissima* showed a strong
153 negative correlation between F_{ST} (q=2) and $N_e m$ estimated from mutual information I (q=1), as
154 expected because F_{ST} and I are each inversely related to $N_e m$ [49]. Dispersal can also be assessed
155 using I in clonal species [50], haploids, and for other ploidies such as a three-species hybrid moss
156 *Sphagnum x falcatulurm* [45].

157 [Insert Fig 1 about here]

158

159 Other equations predicting equilibrium I are based on rigorous theory, rather than simulations,
160 but require knowledge of mutation rate. An equation for bi-allelic SNPs [36] (Box S1) fits closely to
161 simulation outcomes, as well as to data from laboratory colonies of *Drosophila* of known $N_e m$
162 (Figure 1B). Also, equations for IAM and SMM loci [37] (Box S1) agree with simulation outcomes,

and with observed genetic divergence for SMM (microsatellite) loci in starlings (*Sturnus vulgaris*) introduced to Australia [37]. There are many other recent examples of the use of molecular mutual information I to investigate hypotheses about temporal or spatial environmental gradients or barriers, in a variety of free-living and parasitic plants, animals, and fungi, plus simulations (Box S3). In most cases, different *q*-values show similar results, but we discuss divergent cases later. Calculations, sampling and programs, are in Box S6.

Using Information Methods to Scan the Genome for Adaptive Innovation

There is a boom in searches for adaptive genomic regions, for evolutionary interest, choosing reintroduction sources for conservation [51] and identifying human disease loci [52, 53]. Strategies to detect these regions include 'evolve-resequence' or 'transplant' experiments [54], and assessing genomic differentiation over landscapes ('landscape genomics') [55]. Selection acts via differential fitness of variants of any heritable characteristic: DNA sequence variation, expression variability due to interaction between environment and multiple genes, or non-genetic inheritance such as epigenetics [56, 57]. Partly because of these many modes, there are many signals of selection, but each has a high rate of false-positives [58], which can be minimised by using multiple alternative approaches, including (*q*=1) methods.

For detecting selection, Shannon-based metrics are particularly appropriate, in ways such as their greater sensitivity (than *q*=2) to rare alleles that may be important for conservation management or detecting new (potentially adaptive) alleles [59]. For directional selection, which favours a single

advantageous allele [5], Shannon information is proposed as a natural measure of evolvability [38, 39]. The ultimate result will be a single allele at 100%, so ${}^1H$ tends to zero, with dynamics analysed by logit transformation that is algebraically related to mutual information between the allele distributions, before and after selection [39, 60]. (Box S4). Unlike directional selection, balancing selection maintains equilibrium proportions of two or more alleles, for example due to high heterozygote fitness [5], and the predicted allele proportions can be used to calculate the equilibrium value of Shannon information ${}^1H$ (Box S4) [35]. In addition, much adaptive variation is based on quantitative traits controlled by multiple loci. Information approaches to directional selection on multilocus traits (Box S4 [61-64]) indicate that such evolution favours gene duplication [65]. The multilocus analog of balancing selection is called stabilising selection, for which there are also treatments based on information statistics [66].

Detecting when Selection Changes Patterns in either Populations or Expression Profiles

For diploids, departure from single-locus Hardy-Weinberg expectations (HWE) due to selection, non-random mating and other forces is summarised by F_{IS} in (q=2) form (Box S4) [4, 5]. Equivalents for (q=1) include the conventional log-linear-chisquare for fit to random-mating expectations (HWE Box S4), as well as extensions to mixed mating systems, such as mixed selfing and outcrossing, in plants and invertebrates (Boxes S1, S3, S6.3) [12, 34, 67].

Analysis of selection, and other evolutionary processes, requires us to evaluate 'haplotypes' of variants in multiple parts of the genome, including SNPs, insertions, deletions [40]. Discriminating

the individual and combined effects of haplotype elements requires analysis of 'Linkage disequilibrium' (LD) which is non-random association among such variants, due to physical linkage ('true LD' [68]), historical population size and admixture [69], or 'epistatic' selection in which environmental conditions favour particular combinations of variants at multiple positions [7, 52, 53, 70]. LD can be expressed with Mutual Information I (q=1) for association between variants at different positions, which minimises false-discovery of LD (Boxes S5, S6.3) [7, 47, 52, 53, 70-72] and can analyse multi-locus associations of both continuous normal [52] or non-normal [71] traits. This approach has identified the combined effect of SNPs in two protein-coding loci, upon an addictive phenotype [73]. Also, I can be used to assess whether adaptive changes have resulted in different haplotypes in different locations (Box S5) [47, 74, 75]. Physical linkage is not the only way genes interact, and information methods are used to integrate the potentially selective effects of expression networks, DNA sequence frequencies, and linkage disequilibrium [7, 70, 76, 77], for example helping us predict phenotypic outcomes of RNA modification (splicing) in blood-clotting factors [78]. A partial profile ($q = 1,2$) proved best for assessing expression patterns [79]. Programs for analyses are in Box S6.3.

Choosing Measures – the Relative Merits of Information-based Measures and Other Measures.

A full diversity profile exploits the sensitivities, and strengths, of each element (q = 0, 1, 2). Reassuringly, they often give similar results (Box S3), but to understand cases where they differ, or to choose measures for particular applications, we must assess the sensitivities, strengths and weaknesses of each profile element (q=0,1,2). For example, sensitivity to rare alleles may be an

advantage for conservation or adaptation studies (Box 3), but this sensitivity means that missing rare alleles can seriously degrade (q=0) estimates unless appropriate corrections are made (Box S6.2). The performance of measures must be evaluated for each evolutionary or ecological question, such as change of molecular diversity with latitude, and estimation of dispersal from genetic markers (where mutual information outperforms other methods [12]). Here we point out some major differences between (q=1) *versus* (q=0) or (q=2).

Shannon results generally accord with biological predictions, but in some cases the full q-profile aids detection (or rejection) of a predicted pattern, because the (q=1) measure and other measures diverge. Sometimes (q=1) has greater sensitivity than do (q=0) or (q=2) measures (Box S3), as was seen in two studies of alpha (within-population) diversity due to adaptation and drift. In an invasive mosquito *Aedes japonicus japonicus* [80], and in crop carrots affecting nearby wild carrots *Daucus carota* [81], it was found that ${}^{2}H$ (heterozygosity q=2) was less sensitive than ${}^{1}H$ (Shannon q=1) to the predicted loss of variability resulting from small population size and colonisation. This is presumably because ${}^{2}H$ emphasizes common alleles that have relatively low chance of being lost during a bottleneck that was caused by either massive mortality or small numbers of invaders. A hierarchical partition of Shannon information (similar to Box 2-IV) in three-species hybrid moss, *Sphagnum x falcatulurm,* revealed differentiation due to rare recent mutations to which (q=2) metrics would not have been sensitive [45]. However, (q=1) does not always show the greatest sensitivity, in temporal studies of zooplankton, (q=0) was more powerful than (q=1), which was more powerful than (q=2) [28]. Examples in this paragraph demonstrate the importance of analysing the entire q-profile including (q=1).

While recognising the importance of the entire *q*-profile, Table 1 and Boxes 3 and 4 show that only (*q*=1) measures combine a number of desirable characteristics for tracking evolutionarily important phenomena. Firstly, Shannon measures are sensitive to alleles according to their frequency, yet have minimal sampling problems (Table 1, Boxes 1,3). Secondly, measures of each *q*-order must be able to distinguish levels of diversity that are individually important for evolution and conservation [19, 20]: within-locality (alpha), among-locality differentiation (beta), and total (gamma). Eliminating dependencies between these levels for many common (*q*=2) methods requires more complex equations or algorithms [42], which might complicate predictions. In contrast, the explicit hierarchical nesting of information measures (Figure 6 in [10]), yields independent estimates of within-population (α), among-population (β), and total (γ) levels of diversity (Table 1, Box 4). Disagreements between beta *q*=1 and *q*=2 measures may sometimes derive from these dependencies ([81, 82]. Thirdly, some (*q*=1) measures respond in an intuitively appealing fashion to addition of new alleles, behaving in a predictable way with the number of unshared alleles, and satisfying the principles of strong monotonicity and replication (Table 1, Box 4). 'Replication' [15, 83], means that a measure increases linearly when equally diverse and completely distinct groups are pooled in equal proportions, as seen in effective numbers (^{q}D, Boxes 1,S1). Information measures also have minimal sampling problems if appropriate estimation methods are used (Box S6.2). In contrast, a drawback of (*q*=0) measures is that although they can be forecast under specified conditions [5], the sampling to test these forecasts can be hampered by their extreme sensitivity to rare alleles (Boxes 1,3,4,S1,S2).

[Insert Table 1 and Boxes 3 and 4 about here]

Additionally, we have shown above that Shannon measures are catching up in areas where they have lagged behind (q=0) and (q=2): forecasting methods are now available for (q=1) measures of neutral and adaptive variation, and the frequency of use of (q=1) methods in molecular ecology is rising to meet their already heavy use in community ecology. Also, we have shown that Shannon (q=1) methods outperform others for genetic estimation of dispersal [12], and have great utility in detecting selection and gene expression patterns [75, 76]. Thus we should use information methods as important contributors to the diversity profile.

The Near Future: Challenges, Opportunities and Integration

[Box 5 Outstanding questions as sidebar about here]

The major components of molecular information theory are now established, but outstanding questions remain (Box 5). In the long term, the most useful measure will undoubtedly be the whole ${}^{q}D$ profile (${}^{0}D$, ${}^{1}D$, ${}^{2}D$) at α, β, and γ levels. This will maximise understanding of patterns, and allow meta-analysis of the q-profile's performance under a wide variety of conditions.

To complete an Information-based strategy for forecasting and analysis in molecular ecology and evolution, Shannon's strong performance with dispersal, mutation and drift (above) must be extended to include asymmetric dispersal, unequal or changing population sizes, and variants with mutation not described by models such as infinite (IAM) stepwise (SMM) or biallelic (SNP). This could be through new theory, or *via* Approximate Bayesian Computation using the q-profile [12, 27]. It would also be ideal to develop an analogue to AMOVA [44], based on information-theoretic

292 methods. This might build upon a molecular phylogenetic differentiation measure for all *q* [84],
293 based on the neutral coalescent of evolutionary biology [4].

294

295 As well as neutral predictions, we are only beginning to explore the capacity of information science
296 to integrate all analyses of adaptation and selection into a common scale, possibly exploiting the
297 connection to logit (log-linear) modeling. Moreover, sensitivity of tests for loci under selection is
298 likely to increase with a (*q*=1) approach, because they are more sensitive to novel rare variants that
299 are crucial in adaptation [59], and because many of these tests rely upon partition of among-
300 population (beta) *versus* within-population (alpha) measures of diversity [54, 55, 85], and so should
301 benefit from the complete independence of α and β in the (*q*=1) scale.

302

303 Understanding adaptation requires integrating all aspects of biology from sub-cellular biology to
304 habitat tolerance, and information theory is particularly well suited to this challenge [6, 8, 9], being
305 both explicitly additive and hierarchical, as well as being a general forecasting method [1], already
306 used in community ecology [2]. Community ecologists have borrowed genetic (*q*=2) theory for both
307 neutral [13] and adaptive genes [21], so community ecology might also benefit from deploying the
308 (*q*=1) neutral and selective theory outlined in this article. They will likely be able to make even
309 more forecasts than with (*q*=2), by virtue of the fact that Shannon applies to all types of
310 information [86], including physical habitat, behaviour [87], genetics [12, 35-37], information within
311 each sequence including codon usage biases [88], and various non-genetic inheritance modalities
312 including epigenetics [56, 57]. Moreover, information has been proposed as the driver of any
313 correlations between species and molecular diversity [89]. Incorporating information on genetic
314 and/or functional similarity or difference of alleles (or species) greatly improves the utility of

315 diversity measures [83, 90], and ongoing attempts to incorporate function, without violating
316 fundamental diversity principles, should include Shannon methods [83, 91].

317

318 We have only presented a limited aspect of information theory, but we could paint on a much
319 broader canvas. Other potentially fruitful aspects of information theory include: Kullback-Leibler or
320 Relative entropy used in selection analysis; 'Fisher information' used to compare (q=2) measures
321 [92]; fractals [93]; plus compression algorithms proposed for both selection analysis [63] and for
322 joint estimation of sequence alignments and phylogenies [94]. Finally, predictive equations for
323 molecular Shannon information might be used to advance evolutionary computing (machine
324 learning), in which code's performance is optimised by a process similar to biological drift and
325 selection [76, 87, 95, 96]

326

327 Conclusions

328 Four common processes - Dispersal, Random drift, Adaptation, and Generation of novelty
329 (speciation, mutation, recombination) – unify community and molecular ecology [13,21] . The
330 unification of these fields will be accelerated by cross-fertilisation between their predictive
331 equations and measures, at levels from bioinformatics to physical landscapes and beyond,
332 especially when using a diversity profile of q=0,1,2, and adhering to criteria for robust measures
333 (see "Choosing Measures …" section). Shannon Information measures (q=1) are a vital part of this
334 profile, now have predictive equations available, and excel at providing intuitive results.

Supplementary Material

Prediction, Sampling, Estimation, and Examples - abbreviated titles

Box S1 Measuring and Predicting Molecular Entropies and Diversities for (q=1,2,3), Within-populations (alpha) and Between-populations (beta)

Box S2 Derivation of the q-profile

Box S3 Studies using Molecular Information Measures plus q-profile

Box S4 Selection

Box S5 Selection and Linkage Disequilibrium (LD)

Box S6.1 Molecular Information Analyses – Example Calculations

Box S6.2 Molecular Information Analyses – Sampling Considerations

Box S6.3 Molecular Information Analyses – Estimation Programs

Box S7 References

Acknowledgements

We thank reviewers and colleagues who read drafts and made valuable comments: Bragg, Brandenburger, Byrnes, Carmel, Chabanne, Daly, Jabot, Kidd, Kopps, Kosman, Leinster, Manlik, Maslen, Nichols, O'Reilly, Peakall, Raphael, Reeve, Rollins, Shofner, Stear, Sunnucks, Tanaka. TREE editor Craze also made extremely valuable suggestions. WS, AC and LJ first met in 2012 at a workshop coordinated by Leinster Reeve *et al*., on "The Mathematics of Biodiversity" at the Centre de Recerca Matemàtica, Universitat Autònoma de Barcelona, funded by that centre, the Spanish government, and the BBSRC (International Workshop Grant BB/J020567/1).

References

1. Cover, T.M. and Thomas, J.A. (1991) Elements of information theory Wiley.
2. Elith, J. et al. (2011) A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17, 43–57.
3. Bessa, R.J. et al. (2009) Entropy and correntropy against minimum square error in offline and online three-day ahead wind power forecasting. IEEE Transactions on Power Systems 24 (4), 1657-1666.
4. Nielsen, R. and Slatkin, M. (2013) An introduction to population genetics theory and applications, Sinauer.
5. Halliburton, R. (2004) Introduction to population genetics, Pearson.
6. Searls, D.B. (2010) The roots of bioinformatics. PLoS Computational Biology 6 (6).
7. Moore, J.H. and Hu, T. (2015) Epistasis analysis using information theory. In Epistasis: methods and protocols, methods in molecular biology (Moore, J.H. and Williams, S.M. eds).
8. Glazebrook, J.F. and Wallace, R. (2012) 'The frozen accident' as an evolutionary adaptation: a rate distortion theory perspective on the dynamics and symmetries of genetic coding mechanisms. Informatica 36, 53–73.
9. Skene, K.R. (2015) Life's a gas: a thermodynamic theory of biological evolution. Entropy 17, 5522-5548.
10. Shannon, C.E. (1948) A mathematical theory of communication. The Bell System Technical Journal 27, 379–423, 623–656.
11. Buddle, C.M. et al. (2004) The importance and use of taxon sampling curves for comparative biodiversity research with forest arthropod assemblages. Canadian Entomologist 137:120-127.
12. Sherwin, W.B. et al. (2006) Measurement of biological information with applications from genes to landscapes. Molecular Ecology 15, 2857-2869.
13. Hubbell, S.P. (2001) The unified neutral theory of biodiversity and biogeography, Princeton University Press.
14. Jost, L. (2006) Entropy and diversity. Oikos 113, 363–375.
15. Chao, A. et al. (2014) Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. Annual Review of Ecology, Evolution, and Systematics 45, 297-324.
16. Gregorius, H.R. and Kosman, E. (2016) On the notion of dispersion: from dispersion to diversity. Methods in Ecology and Evolution, DOI: 10.1111/2041-210X.12665.
17. Pielou, E.C. (1966) The measurement of diversity in different types of biological collections. Journal of Theoretical Biology 13, 131-144.
18. Hill, M.O. (1973) Diversityand evenness: a unifying notation and its consequences. Ecology 54, 427-432.
19. Anderson, M. et al. (2011) Naviagating the multiple meanings of beta diversity: a roadmap for the practising ecologist. Ecology Letters 14, 19-28.
20. Socolar, J.B. et al. (2016) How should beta-diversity inform biodiversity conservation? Trends in Ecology & Evolution 31, 67-80.
21. Vellend, M. (2016) The Theory of Ecological Communities (MPB-57), Princeton University Press.
22. Luikart, G. et al. (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. Journal of Heredity 89 (3), 238-247.
23. Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585-595.
24. Fu, Y.-X., Li, W-H. (1993) Statistical tests of neutrality of mutations. Genetics 133, 693-709.
25. Greenbaum, G. et al. (2014) Allelic richness following population founding events – a stochastic modeling framework incorporating gene flow and genetic drift. PLoS ONE 9 (12), e115203.
26. Lloyd, M.W. et al. (2013) The power to detect recent fragmentation events using genetic differentiation methods. PLoS ONE 8 (4), e63981.
27. Huang, W. et al. (2011) Approximate Bayesian comparative phylogeographic inference from multiple taxa and multiple loci with rate heterogeneity. BMC Bioinformatics 12, 1.

28. Iacchei, M. et al. (2017) It's about time: Insights into temporal genetic patterns in oceanic zooplankton from biodiversity indices. Limnol. Oceanogr. 00, 00-00.
29. Jost, L. (2008) Gst and its relatives do not measure differentiation. Mol. Ecol. 17, 4015-4026.
30. Crow, J.F. (2001) Shannon's brief foray into genetics. Genetics 159, 915-917.
31. Ewens, W.J. (1972) The sampling theory of selectively neutral alleles. Theoretical Population Biology 3, 87-112.
32. Lewontin, R.C. (1972) The apportionment of human diversity. Evolutionary Biology 6, 381-398.
33. Brown, A.H.D. and Weir, B.S. (1983) Measuring genetic variability in plant populations. In Isozymes in Plant Genetics and Breeding (Tanksley, S.D. and Orton, T.J. eds), pp. 219–239, Elsevier Science Publications.
34. Meirmans, P.G. and Van Tienderen, P.H. (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Molecular Ecology Notes 4, 792–794.
35. Sherwin, W.B. (2010) Review: entropy and information approaches to genetic diversity and its expression: genomic geography. Entropy 12, 1765-1798.
36. Dewar, R.C. et al. (2011) Predictions of single-nucleotide polymorphism differentiation between two populations in terms of mutual information. Molecular Ecology 20, 3156–3166.
37. Chao, A. et al. (2015) Expected Shannon entropy and Shannon differentiation between subpopulations for neutral genes under the finite island model. PLoS ONE 10 (6), e0125471.
38. Day, T. (2015) Information entropy as a measure of genetic diversity and evolvability in colonization. Molecular Ecology 24, 2073–2083.
39. Frank, S.A. (2017) Universal expressions of population change by the Price equation: natural selection, information, and maximum entropy production. Ecology and Evolution 7 (10), 3381–3396.
40. Chin, S.-C. et al. (2016) Phased diploid genome assembly with single molecule real-time sequencing bioRxiv, dx.doi.org/10.1101/056887.
41. Rossetto, M. et al. (2008) Dispersal limitations, rather than bottlenecks or habitat specificity, can restrict the distribution of rare and endemic rainforest trees. . American Journal of Botany 95, 321–329.
42. Hedrick, P.W. (2005) A standardized genetic differentiation measure. Evolution 59 (8), 1633-1638.
43. Pritchard, J.K. et al. (2000) Inference of population structure using multilocus genotype data. Genetics 155, 945–959.
44. Excoffier, L. et al. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131 (2), 479-491.
45. Karlin, E.F. and Smouse, P.E. (2017) Allo-triploid S*phagnum x falcatulum*: single individuals contain most of the Holantarctic diversity for ancestrally indicative markers. Ann. Botany In Press dx.doi.org/doi:10.7282/T3FX7CRH.
46. Smouse, P.E. et al. (2015) An informational diversity analysis framework, illustrated with sexually deceptive orchids in early stages of speciation. Molecular Ecology Resources 15, 1375-1384.
47. Smouse, P.E. and Ward, R.H. (1978) A comparison of the genetic infrastructure of the Yecuana and the Yanomama: a likelihood analysis of genotypic variation among populations. Genetics 33, 611-631.
48. Sokal, R.R. and Rohlf, F.J. (1995) Biometry: the principles and practice of statistics in biological research., Freeman.
49. Rossetto, M. et al. (2011) The impact of distance and a shifting temperature gradient on genetic connectivity across a heterogeneous landscape. BMC Evolutionary Biology 11, 126.
50. Karlin, E.F. et al. (2011) One haploid parent contributes 100% of the gene pool for a widespread species in northwest North America. Molecular Ecology 20, 753-767
51. Houde, A.L.S. (2016) Restoring species through reintroductions: strategies for source population selection. Restoration Ecology 23, 746–753.
52. Chanda, P. et al. (2009) Information-theoretic gene-gene and gene-environment interaction analysis of quantitative traits. BMC Genomics 10, 509.
53. Chanda, P. et al. (2008) Ambience: A novel approach and efficient algorithm for identifying informative genetic and environmental associations with complex phenotypes. Genetics 180, 1191-1210.
54. Schlötterer, C. et al. (2015) Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. Heredity 114, 431-440.

55. Vandepitte, K. et al. (2014) Rapid genetic adaptation precedes the spread of an exotic plant species. Molecular Ecology 23, 2157-2164.
56. Danchin, E. et al. (2011) Beyond DNA: integrating inclusive inheritance into an extended theory of evolution. Nature Reviews Genetics 12, 475-486.
57. Bonduriansky, R. (2012) Rethinking heredity, again. Trends in Ecology & Evolution 27, 330-336.
58. Jensen, J.D. et al. (2005) Distinguishing between selective sweeps and demography using DNA polymorphism data. Genetics 170, 1401–1410.
59. Rollins, L.A. et al. (2016) Selection on mitochondrial variants occurs between and within individuals in an expanding invasion. Molecular Biology and Evolution 33, 995-1007.
60. Vuong, H.B. et al. (2017) Influences of host community characteristics on *Borrelia burgdorferi* infection prevalence in blacklegged ticks. PLoS ONE 12 (1), e0167810.
61. Nourmohammad, A. et al. (2014) Adaptive evolution of molecular Phenotypes. Curr. Opin. Genet. Dev. 23.
62. Riedel, N. et al. (2015) Multiple-line inference of selection on quantitative traits. Genetics 201, 305-322.
63. Kobayashi, T.J. and Sughiyama, Y. (2015) Fluctuation relations of fitness and information in population dynamics. Physical Review Letters 115, 238102.
64. de Vladar, H.P. and Barton, N.H. (2011) The contribution of statistical physics to evolutionary biology. Trends in Ecology and Evolution 26, 424-432.
65. Saito, N. et al. (2014) Evolution of genetic redundancy: the relevance of complexity in genotype–phenotype mapping. New Journal of Physics 16, 063013
66. Nourmohammad, A. et al. (2013) Evolution of molecular phenotypes under stabilizing selection. Journal of Statistical Mechanics: Theory and Experiment stacks.iop.org/JSTAT/2013/P01012 doi:10.1088/1742-5468/2013/01/P01012 arXiv:1301.3981v1 [q-bio.PE].
67. Kosman, E. (2014) Measuring diversity: from individuals to populations. Eur J Plant Pathol 138, 467–486.
68. Wegmann, D. et al. (2011) Recombination rates in admixed individuals identified by ancestry-based inference. Nature Genetics 43, 847-853.
69. Holleley, C.E. et al. (2014) Testing single-sample estimators of effective population size in genetically structured populations. Conservation Genetics 15, 23-35.
70. Wu, C. et al. (2012) Genetic association studies: an information content perspective. Current Genomics 13, 566-573.
71. Yee, J. et al. (2015) Detecting genetic interactions for quantitative traits using spacing entropy measure. BioMed. Research International 2015, ID-523641.
72. Zhang, L. et al. (2009) A multilocus linkage disequilibrium measure based on mutual information theory and its applications. Genetica 137, 355–364.
73. Isir, A.B. et al. (2016) An information theoretical study of the epistasis between the *CNR1 1359 G/A* polymorphism and the Taq1A and Taq1B *DRD2* polymorphisms: assessing the susceptibility to cannabis addiction in a Turkish population. J. Mol. Neurosci. 58, 456-460.
74. Smouse, P. (1974) Likelihood analysis of recombinational disequilibrium in multiple locus gametic frequencies. Genetics 76, 557-565.
75. Yoder, J.B. et al. (2014) Genomic signature of adaptation to climate in *Medicago truncatula*. Genetics 196, 1263–1275.
76. Moore, J.H. et al. (2017) Grid-based stochastic search for hierarchical gene-gene interactions in population-based genetic studies of common human diseases. BioData Mining 10, 19.
77. Wang, S. et al. (2017) Integrative information theoretic network analysis for genome-wide association study of aspirin exacerbated respiratory disease in Korean population. BMC Medical Genomics 10(Suppl 1), 31.
78. von Kodolitsch, Y. et al. (2006) Predicting severity of haemophilia A and B splicing mutations by information analysis. Haemophilia 12, 258-262.
79. Wang, K. et al. (2014) Differential Shannon entropy and differential coefficient of variation: alternatives and augmentations to differential expression in the search for disease-related genes. Int. J. Computational Biology and Drug Design 7, 183.

80. Egizi, A. and Fonseca, D.M. (2015) Ecological limits can obscure expansion history: patterns of genetic diversity in a temperate mosquito in Hawaii. Biological Invasions 17, 123-132.
81. Mandel, J.R. et al. (2016) Patterns of gene flow between crop and wild carrot, *Daucus carota* (Apiaceae) in the United States. PLoS ONE 11 (9), e0161971.
82. Cooke, G.M. et al. (2016) Understanding the spatial scale of genetic connectivity at sea: unique insights from a land fish and a meta-analysis. PLoS ONE 11 (5), e0150991.
83. Leinster, T. and Cobbold, C. (2012) Measuring diversity: the importance of species similarity. Ecology 93, 477-489.
84. Chiu, C.-H. et al. (2014) Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. Ecological Monographs 84, 21–44.
85. Beaumont, M. and Nichols, R.A. (1996) Evaluating loci for use in the geentic analysis of popualtion structure. Proccedings of the Royal Sociey of London 263, 1619-1626.
86. Bergstrom, C.T. and M., L. (2005) The fitness value of information. arXiv q-bio/0510007v1 [q-bio.PE].
87. Dodig-Crnkovica, G. (2017) Nature as a network of morphological infocomputational processes for cognitive agents. Eur. Phys. J. Special Topics 226, 181.
88. Zeeberg, B. (2002) Shannon information theoretic computation of synonymous codon usage biases in coding regions of human and mouse genomes. Genome Research 12, 944-955.
89. Mashayekhi, M. et al. (2014) A machine learning approach to investigate the reasons behind species extinction. Ecological Informatics 20, 58–66.
90. Chao, A. et al. (2010) Phylogenetic diversity measures based on Hill numbers. Philosophical Transactions of the Royal Society B 365, 3599-3609.
91. Scheiner, S.M. et al. (2016) Decomposing functional diversity. Methods in Ecology and Evolution DOI: 10.1111/2041-210X.12696.
92. Verity, R. and Nichols, R.A. (2014) What is genetic differentiation, and how should we measure it—GST, D, neither or both? Molecular Ecology 23, 4216–4225.
93. Zhou, Q. and Yu, Y.-M. (2014) Information dimension analysis of bacterial essential and nonessential genes based on chaos game representation. Journal of Physics D: Applied Physics 47 (46), 465401.
94. Ané, C. and Sanderson, M.J. (2005) Missing the forest for the trees: phylogenetic compression and Its implications for inferring complex evolutionary histories. Syst. Biol. 54, 146-157.
95. Hadka, D. and Reed, P. (2013) Borg: an auto-adaptive many-objective evolutionary computing framework. Evolutionary Computation 21, 231-259.
96. Sole, R. (2016) Synthetic transitions: towards a new synthesis. Phil. Trans. Roy. Soc. B. 371, 2015.0438.
97. Mather, A.E. et al. (2012) An ecological approach to assessing the epidemiology of antimicrobial resistance in animal and human populations. Proc. R. Soc. B. 279, 1630-1639.
98. Jost, L. et al. (2010) Partitioning diversity for conservation analyses. Diversity and Distributions 16, 65–76.
99. Chiu, C.-H. and Chao, A. (2014) Distance-based functional diversity measures and their decomposition: A framework based on Hill numbers. PlosONE 9 (7), e100014.
100. Chao, A. and Chiu, C.-H. (2016) Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures. Methods in Ecology and Evolution 7, 919-928.

**** FIGURE 1 CAPTIONS (NB FIGURES SUPPLIED AS .pptx)*****

Figure 1. Relationship of Mutual Information *I* to Population Size and Dispersal, from Simulations and Living Populations.

(A) Observed Mutual information I per locus from simulated microsatellite data, used to estimate dispersal ($N_e m$) via F_{st} or via I. Root mean square error (RMSE) of $N_e m$ is plotted against dispersal rate, for several different effective population sizes (N_e). RMSE is similar to variance, except it assesses departure of (simulated) observations from the equation's prediction, rather than from the mean of the observations. In every case, $N_e m$ calculated from I had the lower RMSE (from [12]).

(B) Comparison of analytical predictions of mutual information I with observed SNP mutual information in *Drosophila* fly dispersal experiments with known $N_e m$. For the predictions, a mutation rate of $\mu=10^{-6}$ was assumed, but using $\mu=10^{-3}$ to 10^{-9} made little difference to the predicted values of I (modified from [36]).

570 ********************BEGIN TABLE 1 *******************

571 **Table 1. Relative Sensitivities, Strengths, and Weaknesses of Each Element of the Diversity Profile**
572 **From (*q*=0 to 2) (Boxes 1, S1)**; citations are in text.

573

Value of *q*	Interpretation of alpha (between-population) and beta (within-population) measures	Advantages	Disadvantages	Possible fixes for disadvantages
q=0	**Within-population measure $^0H_\alpha$:** Number of different types of alleles (or species) (Box 1).	More sensitive to rare alleles than *q*=1 or *q*=2.	Serious sampling problems, because very sensitive to rare alleles.	Sampling problems somewhat alleviated by method in Box S6.2
	Between-population (*β*): Several measures (Box S1), all related to the number of allelic types that are NOT shared between the localities.	More sensitive to rare alleles than *q*=1 or *q*=2.	Serious sampling problems, because very sensitive to rare alleles.	Sampling problems somewhat alleviated by method in Box S6.2
q=1	**Within-population $^1H_\alpha$:** Number of ways the array of different alleles could be set out, given the relative fractions of different alleles, p_i available (Box 1). More formally, higher ^1H means that there is reduced certainty about what type to expect when a single allele is randomly sampled.	The most commonly used abundance-sensitive measure in community diversity. Sensitivity to each allele according to its frequency, *ie*.: each copy of each allele is treated equally.	Some sampling problems	Sampling problems fixed by method in Box S6.2
	Between-population (*β*): Several measures (BoxS1), all related to whether knowing the allelic type of ONE single individual will help identify the location from which it was sampled (eg, if there is no differentiation, then the allelic information is completely uninformative, but if there is complete	Shannon differentiation satisfies 'monotonicity' (Some other transformations for (*q*=1) do not). Shannon differentiation (Box 2) satisfies 'true dissimilarity', which means that the differentiation	Some sampling problems	Sampling problems fixed by method in Box S6.2

	differentiation, then knowing the allelic type would give certain identification of the location of origin)	measure should represent the actual proportion of non-overlapping alleles, when populations are equally diverse and all alleles have the same frequencies. Some other transformations for (q=1) do not satisfy this. Shannon information provides a natural measure of evolvability.		
q=2	**Within-population $^2H_\alpha$:** Chance of choosing TWO different types, given the relative fractions of different types, p_i available (Box 1).	Very sensitive to frequent types. Relatively few sampling problems	Insensitive to rare types that may be important in conservation and long-term evolution	
q=2	**Between-population (β):** Many measures (Box S1), all related to whether TWO individuals sampled from different localities are likely to have different allelic types (eg if there is no differentiation, they MUST be the same allelic type, but if there is complete differentiation, they MUST be different types).	Very sensitive to frequent types. Relatively few sampling problems *Jost-D* (Box S1) satisfies 'true dissimilarity', which means that the differentiation measure should represent the actual proportion of non-overlapping alleles, when populations are equally diverse and all alleles have the same frequencies. Some other transformations for (q=2) do not satisfy this.	Insensitive to rare types that may be important in conservation and long-term evolution Most (q=2) measures confound among-population (beta) diversity with one or both of within-population (alpha) diversity and total (gamma) diversity. These are cases of the 'dependence-on-alpha problem' and the 'dependence-on-gamma problem'. Most (q=2) measures lack monotonicity, ie apparent difference between two	*Jost-D* (Box S1) fixes the dependence-on-alpha problem. A measure to fix the dependence-on-gamma problem is cited in the main text.

			populations can decrease when a new unshared allele appears in a population, wrongly implying that this reduces the pace of speciation. The q = 2 differentiation measure *Jost-D* (Box S1), does not possess the expected monotonicity. F_{ST} (Box S1), which is also sometimes used as a beta differentiation measure, does not satisfy monotonicity.	

574

575

576

577

***************START BOX 1 **************************

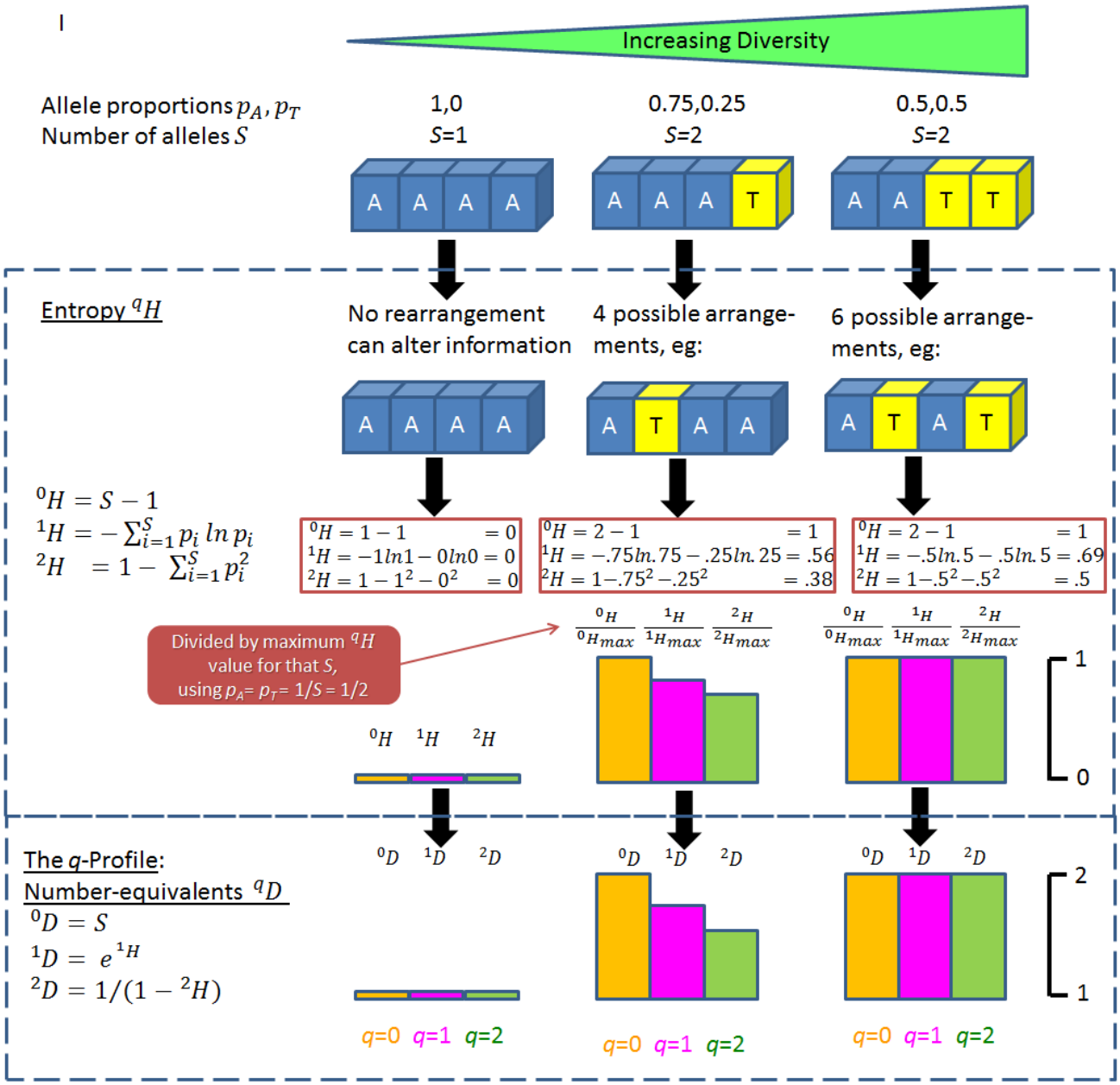
Box 1 Tutorial on the q-Profile: 'Effective-Number' Diversities ${}^0D$, ${}^1D$, ${}^2D$ and their derivation from the corresponding Entropies ${}^0H$ ${}^1H$, ${}^2H$.

Figure I shows three samples of four haploid individuals, each genotyped to identify SNP allele *A* or *T*. The first row of histogram bars shows entropy values ${}^qH$, including heterozygosity ${}^2H$, and ${}^1H$ which can be derived from the number of possible novel arrangements of the alleles carried by the four sampled individuals, as if one was trying to spell out words with the alleles. In the left sample, which has no diversity, all ${}^qH$ measures are zero (NB conventionally, $0\ ln\ 0 = 0$). In the right sample, where alleles are equally-frequent, each measure is at its maximum possible value with two alleles. The second row of histogram bars shows the *q*-profile of 'effective-number' diversities ${}^qD$ derived from the ${}^qH$ entropies. Note that in moving from the left sample to the middle sample, we are adding a rare allele, a single copy of a new allele *T*. In this case, ${}^0H$ and ${}^0D$ show the greatest response, while heterozygosity ${}^2H$ and its transform ${}^2D$ show the smallest response, because in Equation 2 when a rare allele's proportion is squared, its effect becomes much smaller or even negligible. In contrast, when both alleles are already present, ${}^0H$ and ${}^0D$ show no response to changing the numbers of copies of each allele, while ${}^2H$ and ${}^2D$ show the greatest response. In both cases ${}^1H$ and ${}^1D$ show an intermediate response. All formulas are in Box S1.

Box 1 continues on next page…

Box 1 continued.

Figure I. Calculating ${}^{q}H$ and ${}^{q}D$



Box 1 continues on next page…

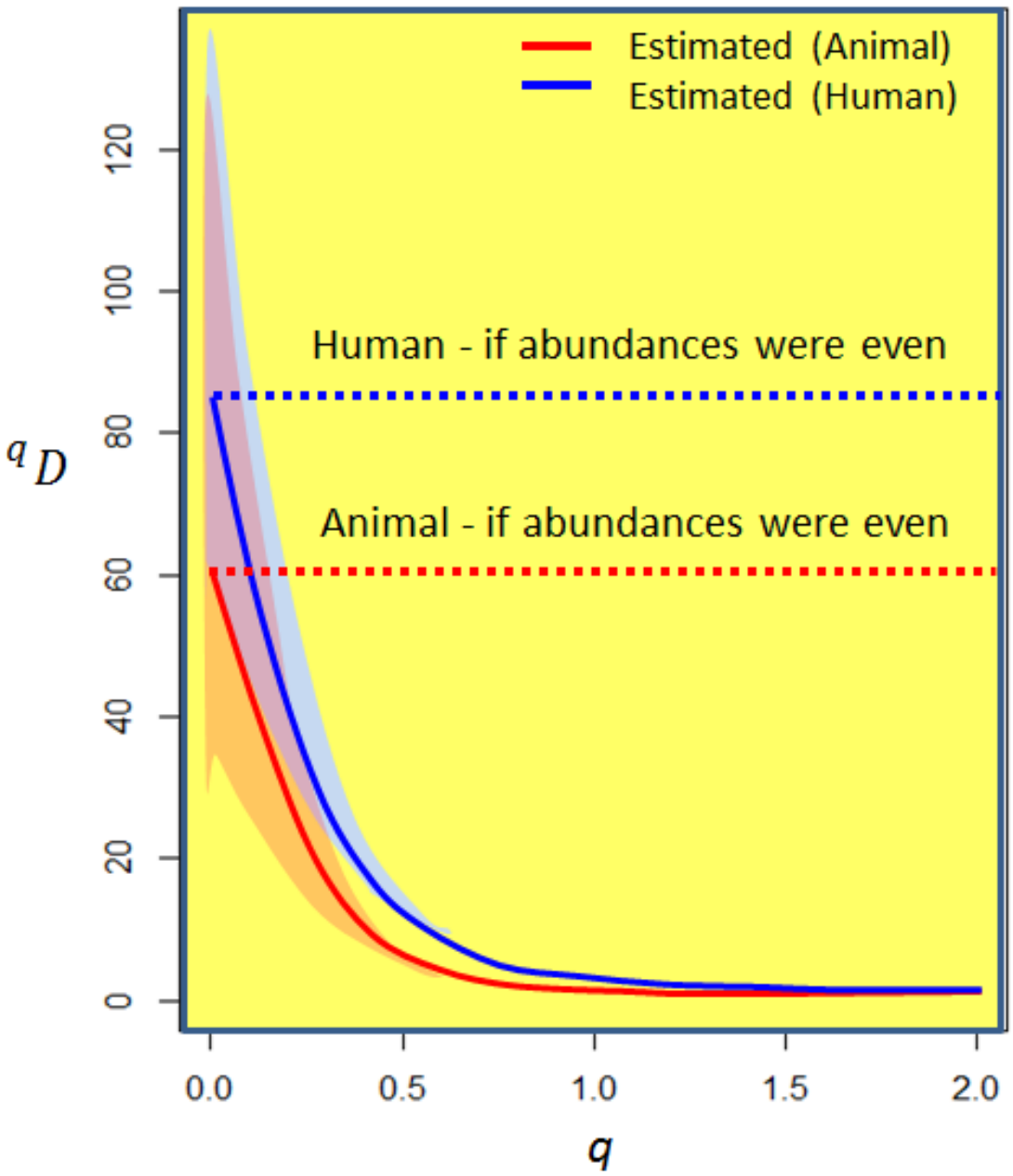
606 Box 1 continued.

607 Figure II shows diversity profiles of antimicrobial resistance variants within *Salmonella typhimurium*
608 in humans (blue) and domestic animals (red) [97], with correction for sampling bias (Box S6.2). The
609 horizontal axis is the value of q, and the vertical axis is ${}^{q}D$ for each value of q. The shaded areas
610 are 95% confidence limits. Note that these limits overlap, except in the region ($0.3 < q < 1.2$) where
611 the plots can be discriminated. This highlights the difficulty of comparing profiles if only (q=0) and
612 (q=2) are used. Also note the effect of evenness of observed allelic distributions upon diversity
613 profile. If the resistance types had been equally frequent within each host ($p_1 = p_2 =$
614 $p_3 = \cdots = p_S$) then the ${}^{q}D$ profiles would have been flat (dotted lines). However, in each group
615 (Animal, Human), there is one highly abundant type and a many types only recorded once.
616 Therefore, the profiles drop very sharply and become flat for order (q=2) (*eg*. Heterozygosity or
617 Gini-Simpson), because for this q–value, ${}^{q}D$ is mainly determined by the abundant type(s), being
618 insensitive to rare types.

619 Box 1 continues on next page…

620 Box 1 continued.

621 Figure II q-Profiles

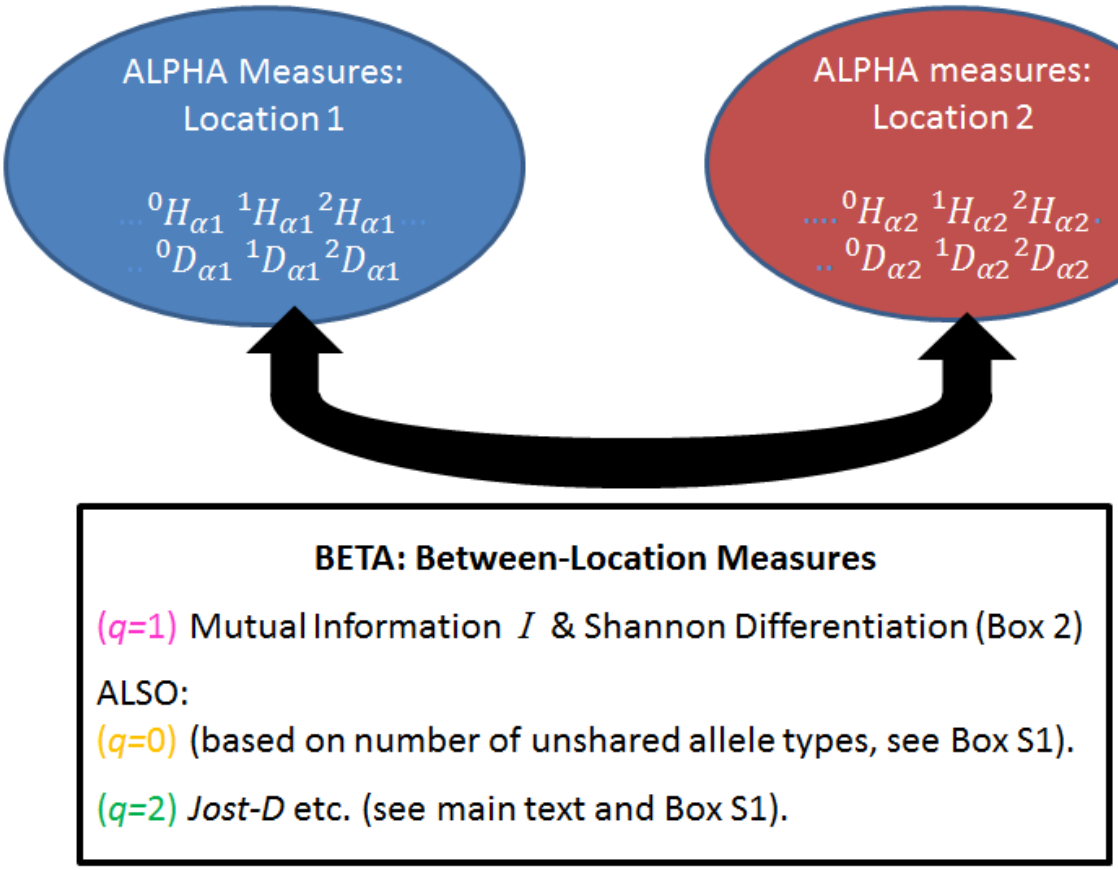


622
623
624 Box 1 continues on next page…

625 Box 1 continued.

626 **Figure III Alpha and Beta Measures of Entropy and Diversity**

627



628

629

630

631

632 ****************END BOX 1 ***************************

**********START BOX 2 **********

Box 2 Information Measures of Geographic Differentiation: Mutual Information I.

'Mutual information' I is a (q=1) measure that has a similar role to correlation in the (q=2) scale; it expresses association between two variables, such as allelic differentiation and population membership ([12], where I is called ${}^{S}H_{UA}$). Therefore differentiation among populations can be measured as mutual information I between geographic location and allelic differentiation. For example, we ask: "Does knowing the alleles in a sample, give any information about which location was sampled?" This is not true in Figure I, where the allele proportions p_C and p_T are the same in the two locations, so there is zero mutual information between the variables 'location' and 'allele type' – in other words allelic data provide no information about population of origin. However in Figure II, mutual information is maximal: knowing the alleles in a sample gives perfect information about which location was sampled, because there are no alleles that are shared between the two locations.

Figure I. Zero Mutual Information I.

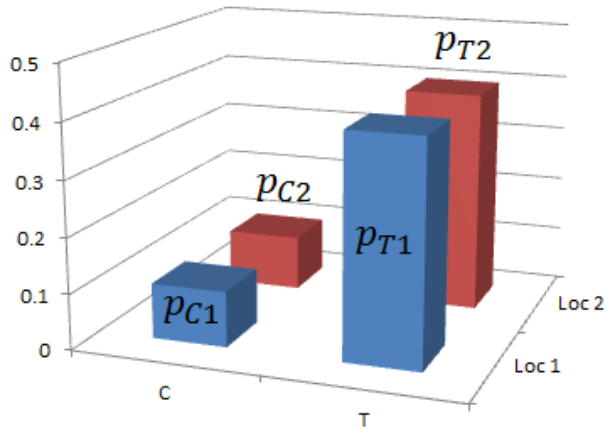
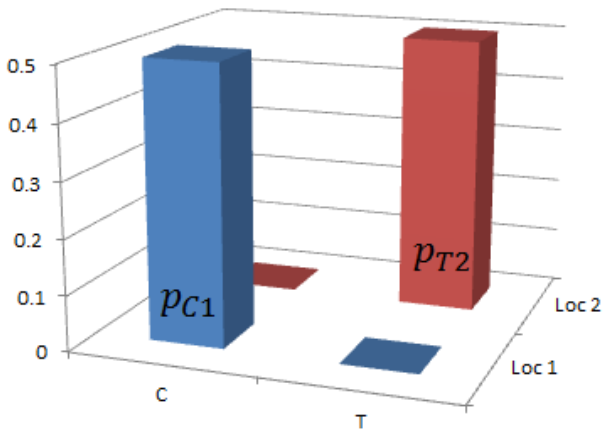


Figure II. Maximum Mutual Information I.



Box 1 continues on next page…

Box 1 continued.

Figure III outlines the calculation of mutual information I from observed data for two locations (*eg,* alpha-level variation within estuary 1, and within estuary 2) as well as total for both locations (gamma-level proportions, marginals p_C and p_T). Mutual information can be derived very easily from the chisquare for allelic differentiation between populations, using $I = \chi^2/2\,n$, (where *n* is the total sample size, I is the mutual information, and χ^2 is log-linear-chisquare, with expectations for each cell calculated as shown for one example in Figure III [48]).

Alternatively, I is calculated as the part of the total information ${}^1H_\gamma$ that is not due to variability within single locations ${}^1H_\alpha$; using terms from Figure III:

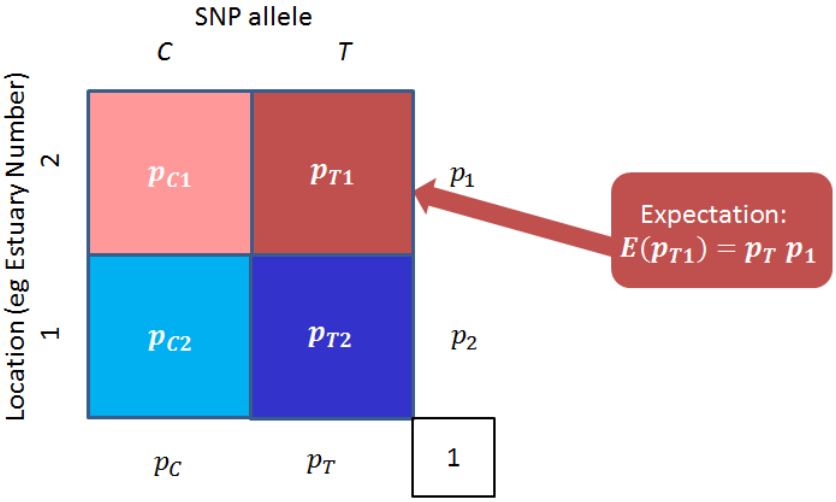
$$I = \left({}^1H_\gamma - \overline{{}^1H_\alpha}\right)$$

$$= \left(-p_C \ln p_C - p_T ln\, p_T\right) - 0.5\left(-\frac{p_{C1}}{p_1}\ln\frac{p_{C1}}{p_1} - \frac{p_{T1}}{p_1}\ln\frac{p_{T1}}{p_1}\right) - 0.5\left(-\frac{p_{C2}}{p_2}\ln\frac{p_{C2}}{p_2} - \frac{p_{T2}}{p_2}\ln\frac{p_{T2}}{p_2}\right)$$ (Box S1).

Also, mutual information adjusted to range from zero to unity is:

Shannon Differentiation = $I/\ln K$ (where *K* is the number of equal-sized populations; Box S1 shows the formula for other cases). Shannon differentiation has useful properties, discussed in Box 4.

Figure III. Data for Calculating Mutual Information between Allele-type and Location.

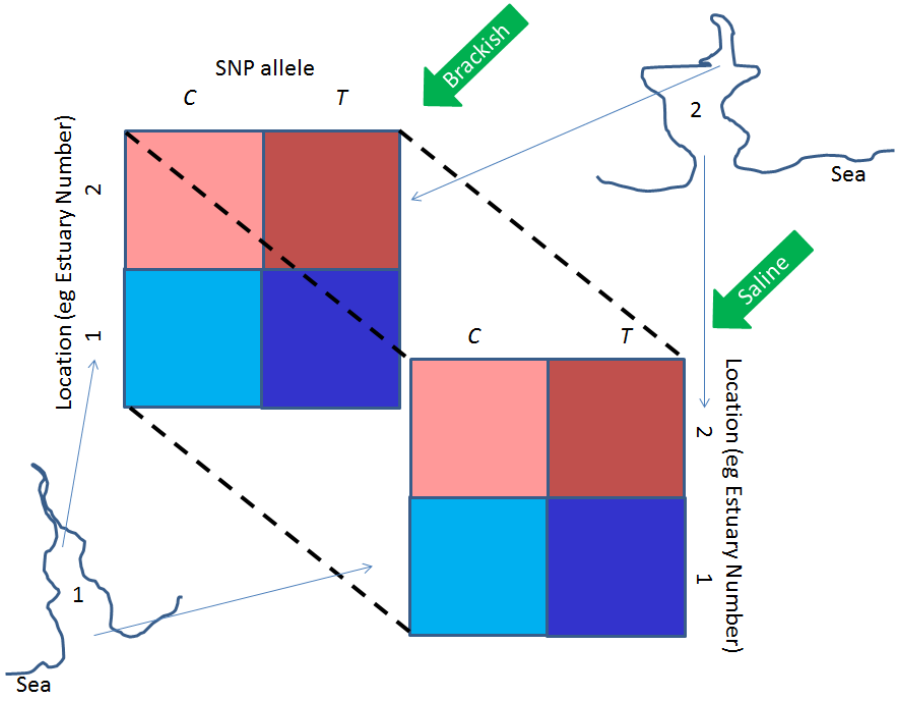


Box 2 continues on next page….

Box 2 continued….

Figure IV shows expansion to include two adjacent habitats (brackish and saline) sampled within each estuary. Each estuary has a saline habitat near the mouth, whose data from Figure III are shown in the foreground of the cube, and a similar set of data for a brackish habitat further inland are in the background of the cube [35]. This expansion for multiple variables is standard for contingency-analysis [48], so the equivalent of mutual information can easily be calculated for this situation [35]. In fact, indefinite expansion is possible, to accommodate multiple alleles per locus, plus dimensions added to incorporate diversity within and among different habitats, landscapes, etc. This is possible because log-linear χ^2 (and therefore I) are completely additive [35, 48]. The partition strategy will depend upon the hypothesis being tested [45, 46, 48]. Programs, sampling bias corrections, and example calculations are in Box S6.

Figure IV. Partitioning Molecular Information with Two Variables (Location and Habitat).



Box 2 continues on next page….

682 Box 2 contined….

683 With equal population sizes, I is the ecologists' Horn measure [37]. Mutual information is closely
684 related to the relative entropy (Kullback-Leibler) which is used for tests for Hardy-Weinberg
685 equilibrium (Box S4) and for comparing allele frequencies before *vs*. after selection, as well as the
686 rate of change of Fisher information [92].

687 ****************END BOX 2 ***************************

688 ********************BEGIN BOX 3 *********************

689 **Box 3 Choosing a Measure for Within-population Diversity (alpha):**

690 **the Relative Merits of Information-based Measures (*q*=1) and Other Measures.**

691 Elements of the diversity profile from (*q*=0) to (*q*=2), have various strengths and weaknesses, each
692 being sensitive to particular aspects of diversity relevant to different questions (Boxes 1,S1).
693 Measures based on counts of different types of alleles (or species, *q*=0) are very sensitive to rare
694 alleles (Box 1), which is obviously important if there is a focus on rarity, either for conservation
695 reasons, or because novel adaptive mutants are initially rare [4, 5, 59]. On the other hand, the
696 (*q*=2) measures such as heterozygosity (${}^{2}H_{\alpha}$ Equation 2 in main text, also Box 1) give very little
697 weight to rare alleles, because they are based on the chance of choosing two different types, given
698 the proportions p_i available in the population. Estimating heterozygosity involves squaring
699 $p_i$values, so that values close to zero (*ie*. rare alleles) become vanishingly small, relative to more
700 common alleles. The intermediate value of *q* (*q*=1, ${}^{1}H_{\alpha}$ Shannon information, Equation 1 in main
701 text, also Box 1) assesses diversity as the number of ways the array of different alleles could be set
702 out, given the p_i values. Shannon weights each allele by its frequency, so its response to addition of
703 a single novel allele is intermediate between those of (*q*=2) and (*q*=0) measures (Box 1-I).

704

705 These sensitivities to rare alleles result in different sampling properties. Counts of different types
706 (*q*=0) are changed considerably by either adding a single individual of a different type, or else failing
707 to sample this single individual. As a result, even after sampling corrections, these measures often
708 cannot distinguish diversity levels of assemblages that can be discriminated by either (*q*=1) or (*q*=2)
709 scales (Box 1-II in main text). In contrast, (*q*=2) methods such as heterozygosity/Simpson's have

relatively few sampling problems. Sampling for Shannon information (q=1) can be well addressed by modern corrections (Box S6.2), so that it can distinguish between alternative assemblages (Box1-II).

Frequency of use is important for comparability between studies. In community diversity, Shannon is the most commonly used abundance-sensitive measure (q =1 ${}^{1}H_{\alpha}$ [11]), whereas heterozygosity (q=2) is most commonly used in molecular ecology [4, 5]. As the two fields gradually unify [13, 21], it will become important to use measures that are common to both fields, and we suggest that a profile of q = 0, 1, 2 will achieve this best.

Finally, there is extensive literature on neutral and adaptive processes in molecular ecology for both (q=0) and (q=2) alpha-measures [4, 5]. More recently, it has been proposed that the (q=1) measure Shannon information ${}^{1}H_{\alpha}$ provides a natural measure of evolvability [38, 39], and Shannon predictive methods are being developed for both neutral [12, 36, 37] and adaptive variants [35, 61-64, 66].

**********************END BOX 3 **********************

**********BEGIN BOX 4 **********

Box 4 Choosing an Among-Population Differentiation Measure (Diversity).

Among-population measures (*q*=0,1,2) inherit virtues and shortfalls of corresponding alpha measures (Box 3), plus properties specific to the beta level. For three reasons, the (*q*=1) measures might be the best tools to track and understand evolutionary processes of divergence.

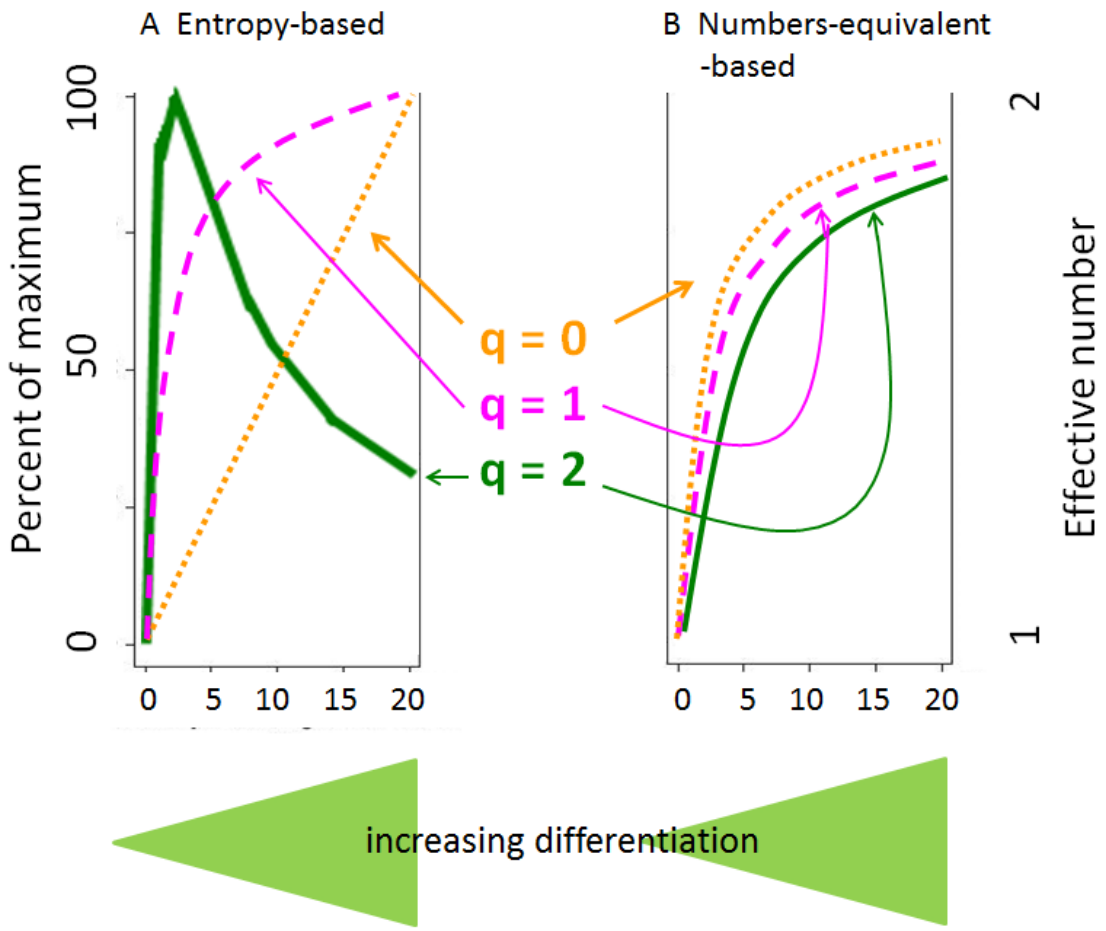
First, many measures used as beta-differentiation measures actually confound this with within-population (alpha) or total (gamma) diversity, *e.g.* F_{ST} and G_{ST} (*q*=2) tend towards zero as diversity increases [98, 99] (Figure I). In contrast, Shannon differentiation (*q* = 1, Box 2) and *Jost-D* (*q*=2 , Box S1) are zero only when allele frequencies are identical across localities, and unity only when there are no shared alleles. Chao and Chiu [100] proposed a measure to avoid dependence-on-gamma.

Box 4 continues on next page….

Box 4 continued.

Figure I Effect of increasing allelic differentiation upon differentiation measures for (*q*=0,1,2).

Initially two populations shared the same two alleles, and the horizontal axis shows the addition of extra unshared alleles to each locality. Equations and symbols are as shown in Box S1, except the *q* =0 entropy-based measure is $S_\gamma - (S_{\alpha 1} + S_{\alpha 2})/2$ and the (*q* =2) entropy-based measure is $\left({}^2H_\gamma - \overline{{}^2H_\alpha}\right) = G_{ST} * {}^2H_\gamma$. In panel A, the vertical axis for each plot is adjusted to percent of the maximum observed differentiation value for that order of *q* (modified from [98]).



Box 4 continues on next page….

Box 4 continued.

Second, to track evolving divergence among localities, a differentiation measure should never decrease when shared alleles are replaced or augmented by new unshared alleles, called 'monotonicity'. Surprisingly, most differentiation measures derived from diversity partitioning do not meet this basic criterion, wrongly implying that new unshared alleles can reduce the pace of differentiation. For example, the additive (q=2) differentiation (Figure I, green solid line, panel A) increases then decreases as differentiation increases, as do most (q=22) measures (e.g., F_{ST} and *Jost-D* Box S1), and even some (q=1) measures. The only differentiation measures with strong monotonicity are (q=1) Shannon differentiation, and those based on (q=0), though the latter are less desirable because of their sampling problems.

A third important property of differentiation measures is 'true dissimilarity', which means that the differentiation measure should represent the actual proportion of non-overlapping alleles, when populations are equally diverse and all alleles have the same frequencies. This is untrue for many (q=1 or 2) measures, but is true for the (q=1) measure Shannon differentiation (Box 2) and for (q=2) differentiation *Jost-D* [29].

**********************END BOX 4 **********************

**********BEGIN BOX 5 OUTSTANDING QUESTIONS *********

Box 5 Outstanding Questions

Regular use of a formal ${}^{q}D$ profile of at least (q = 0, 1, 2) measures at each level (α, β, and γ) will achieve maximum understanding of patterns, and will later allow meta-analysis of the performance of the molecular q-profile under a wide variety of conditions.

Neutral forecasts need to be extended to complex scenarios, such as asymmetric dispersal, unequal population sizes, and bottlenecks including possible separate effects of actual and effective population size.

We need an analogue to AMOVA based on information-theoretic methods.

There is a need for new predictive theory for variants with mutation not described by models such as infinite (IAM) stepwise (SMM) and biallelic SNPs.

Further analyses of adaptation and selection with (q=1) approaches will profit from three attributes of Shannon: similarity to logit (log-linear) modeling, sensitivity to rare novel variants that are crucial in adaptation, and independence of α and β.

Analyses of linkage disequilibrium and expression, which already use measures related to mutual information, might benefit from using its transform to Shannon differentiation.

Box 5 Outstanding Questions continues on next page….

795 Box 5 Outstanding Questions continued.

796 Integrating all biological levels from community ecology and evolution, through to sub-cellular
797 biology, can capitalise on existing protocols for using information and entropy methods as general
798 forecasting methods. This will include incorporating information on genetic and/or functional
799 similarity or difference of alleles (or species).

800

801 Information and entropy theory is very broad, with an abundance of other possible connections
802 within and outside biology.

803 ************END BOX 5 OUTSTANDING QUESTIONS ***********

804 ************BEGIN GLOSSARY BOX ***********

805 ## Glossary

806 **Adaptation:** This refers to the evolutionary process due to natural selection for organisms that are
807 better at surviving and/or reproducing in a particular environment.

808

809 **Additivity**: This refers to a multidimensional table (e.g., dimension 1 is allele type, dimension 2 is
810 location, dimension 3 is an environmental variable, etc). In this type of table, measures such as log-
811 linear contingency chisquare (i.e., mutual information) can be partitioned into completely additive
812 sub-investigations, unlike Pearson's chisquare.

813

814 **Allele proportions and frequencies:** For conformity to conventions in information and entropy
815 theory, statistics, and all other science (except population genetics!), we refer to 'allele-
816 frequencies' when dealing with counts ranging from zero to infinity, and to 'allele-proportions',
817 when these frequencies have been converted to p_i ranging from zero to unity.

818

819 **Alpha, beta, and gamma diversity measures (α, β, γ):** These indicate within-locality (alpha)
820 diversity, among-locality differentiation (beta) diversity, and total (gamma) diversity. In other
821 publications, alpha values are often assumed to be averaged over many locations. Where an
822 average is made, we will indicate this by an overbar, and describe any unequal weighting.

823

824 **Bottleneck:** A period of reduced population size, which usually will alter entropy and diversity
825 levels away from the previous expectations.

826 Glossary Box continues on next page….

827 Glossary Box continued

828 **Drift:** Random genetic drift is caused by the chance nature of transmission of alleles within each
829 family. In a finite population, this chance in transmission results in fluctuations of allele proportions
830 in the entire population. Drift erodes genetic variation summarised by any measure (q=1,2,3).

831

832 **Effective numbers (or 'true diversities') ^{q}D:** These are conversions of entropies (^{q}H) into ^{q}D the
833 number of equally-frequent alleles (or species) that would give the actual observed ^{q}H, derived
834 from a possibly uneven array of alleles in the observed sample.

835

836 **Entropy (^{q}H):** Entropies include Heterozygosity or Gini-Simpson ($^{2}H = H_e$), Shannon (^{1}H), and the
837 number of allelic types minus one (*S*-1, or ^{0}H). The '*H*' symbol is used for entropy throughout
838 science.

839

840 **Epigenetics**: This is one type of non-genetic inheritance, due to chemical modification of DNA
841 sequence, e.g., methylation, which is sometimes transmissible through at least one generation, and
842 may have phenotypic effects. We do not use the term 'epigenetics' to cover all types of non-
843 genetic inheritance. Frequently confused with epistasis.

844

845 **Epistasis**: Functional interactions between multiple loci, at any level from transcription onwards, to
846 affect a measured phenotype. Frequently confused with epigenetics.

847

848 **Haplotype**: A collective genetic combination located on a single molecule of DNA, typically
849 including two or more SNPs showing variation of bases between alternative haplotypes.

850 Glossary Box continues on next page….

Glossary Box continued

Hardy-Weinberg-Equilibrium (HWE): For a single locus, HWE is the condition where the combinations of alleles in diploid genotypes are as expected from random combination of the population's pool of alleles. HWE can be disrupted by mutation, selection, dispersal, non-random mating (including inbreeding), or random genetic drift.

IAM Infinite allele model of mutation: This is suitable for long genomic regions with rare base substitutions, so that each mutation creates an allele that has never existed before. This is approximately suitable for protein-coding regions, and much other non-repetitive DNA.

Information theory: This has multiple strands, but we focus on Shannon's Information Index ${}^{1}H$, which is a measure of the number of different ways that a group of objects (e.g., individuals of different species, or DNA molecules with different sequences) can be rearranged. The information index is also a measure of 'surprise' or uncertainty, increasing in groups where we are less certain of what we would sample at random.

Linkage disequilibrium (LD): This is when the combinations of alleles at two or more diploid loci depart from those expected by random sampling from the population. The loci might be carried on the same molecule of DNA (true LD), or on different molecules (sometimes called genotypic disequilibrium or GD). LD results from random genetic drift, mutation, selection, dispersal, non-random mating (including inbreeding), clonal or asexual reproduction.

Glossary Box continues on next page….

Glossary Box continued

Monotonicity: This means that increases of variable '*x*' are EITHER always associated with an increase of '*y*', OR always associated with a decrease of '*y*'. 'Weak monotonicity' allows plateaus of '*y*'.

Neutral: This refers to genetic variants that do not affect fitness (survival and/or reproduction), so are not involved in natural selection and adaptation.

Non-genetic inheritance: Any type of inheritance that does not rely upon variation of the DNA sequence of A,T,C,G. Examples include microbiome inheritance, epigenetic inheritance, niche inheritance, etc. All of these can have profound fitness effects.

Replication: This means that a diversity measure increases linearly when equally diverse and completely distinct groups are pooled in equal proportions.

Selection: This is based on individuals that possess heritable characteristics (eg alleles) that give them higher relative survival and/or reproduction in a particular environment. As a result, there will be increased representation of those heritable characteristics in the next generation. Such individuals are said to have higher fitness in that environment.

SMM Stepwise mutation model: An approximation of microsatellite evolution.

SNP Single Nucleotide Polymorphism: A single-basepair locus that varies in the population.

************END GLOSSARY BOX ***********