Genomic quantitative genetics to study evolution in the wild

Phillip Gienapp[1], Simone Fior[2], Frédéric Guillaume[3], Jesse R. Lasky[4], Victoria L. Sork[5], Katalin Csilléry[3,6]

[1] Dept. of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, Netherlands

[2] Plant Ecological Genetics, ETH Zurich, Switzerland

[3] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Switzerland

[4] Department of Biology, Pennsylvania State University, USA

[5] Department of Ecology & Evolutionary Biology, University of California Los Angeles, USA

[6] Biodiversity and Conservation Biology, WSL Swiss Federal Research Institute, Birmensdorf, Switzerland

Corresponding author: Gienapp, P. (p.gienapp@nioo.knaw.nl)

Keywords

adaptation, high-throughput genotyping, quantitative genetics, relatedness, natural populations

Abstract

Quantitative genetic theory provides a means of estimating the evolutionary potential of natural populations. However, this approach was previously only feasible in systems where the genetic relatedness between individuals could be inferred from pedigrees or experimental crosses. The genomic revolution opened up the possibility of obtaining the realized proportion of genome shared among individuals in natural populations of virtually any species, which could promise (more) accurate estimates of quantitative genetic parameters in virtually any species. Such a 'genomic' quantitative genetics approach relies on fewer assumptions, offers a greater methodological flexibility, and is thus expected to greatly enhance our understanding of evolution in natural populations, for example, in the context of adaptation to environmental change, eco-evolutionary dynamics, and biodiversity conservation.

Understanding trait evolution in natural populations using quantitative genetics

Reliably predicting responses to selection is essential in animal and plant breeding, for studying evolution in natural populations, and understanding population and species response to rapid anthropogenic environmental change [e.g. 1]. Building on the seminal work of RA Fisher [2], who first proposed to decompose the variation in a continuous trait into genetic and environmental components, breeders and evolutionary biologists have long used quantitative genetic (QG) theory to estimate the heritable proportion of trait variation and co-variation with other traits to predict response to natural and artificial selection [3]. Partitioning the variance in a trait has been possible by assuming a simple model of genetic determinism: each trait is affected by many loci of individually (infinitesimally) small effects, known as the infinitesimal model (see Glossary). Despite this strong, and not necessarily realistic assumption underlying QG [4], it has been very efficient and successful in predicting selection responses in animal and plant breeding over the last century [5].

The key requirement for any QG analyses is information about the relatedness among individuals. The resemblance between relatives is the basis for separating the heritable proportion of the trait variance (the additive genetic variance) that determines the response to selection and potential for evolution. The generalization of QG theory to include any kind of relatives in a single statistical framework called the ‘animal model’ has become the major tool in animal breeding to predict the ‘genetic merit’ or breeding value of individuals [6]. The animal model also opened up the possibility to study evolution in the wild, because it was no longer necessary to perform crosses to obtain many pairs of individuals of the same relationship, instead, any degree of relatedness was regarded as informative. Nevertheless, the application of QG in wild populations remained a tool of privilege, limited to long-term studies, where pedigrees were possible to establish from individual observations, biasing the approach towards a handful of mammal and bird species [7].

With the genomic revolution, QG moved away from the spotlight, and the focus shifted to searching for the genetic determinants of phenotypic variation, i.e. mapping quantitative trait loci (QTL). Thanks to lower genotyping costs, mapping at the finest possible level was enabled by genome wide association studies (GWAS) that aim to find markers, most often single nucleotide polymorphisms (SNPs), associated with phenotypic variation [e.g. 8]. However, GWAS so far did not meet the high expectations: in many cases, a large part of the phenotypic variance remained unexplained by SNPs that reached genome-wide significance (the so-called missing heritability problem); a problem that is now partly resolved by increased sample sizes [8]. Furthermore, even the largest GWAS studies in humans have not yet reached a plateau of the number of 'risk factors' (significant SNPs) associated with trait variation [8]. Not so surprisingly, GWAS conducted in natural plant and animal populations had so far limited power and often failed to identify SNPs associated with traits [reviewed in 9]; yet the lack of any genome-wide significant loci indicates the absence of major effect loci. Some of these studies instead found that the proportion of variance explained by a chromosome correlated with its size [10, 11], as expected if these traits are under the influence of many loci of small effects. All these findings support that the infinitesimal model of classical QG

theory is a reasonable approximation for most quantitative traits, including life-history, behavior, or morphology [12]. The QG approach also has the advantage of not relying on unknown underlying molecular or functional details of the genetic basis of quantitative traits and directly provides estimates of the evolutionary potential of such traits [13]. In this Opinion paper, we promote a so-called 'genomic QG', which makes use of genomic data almost exclusively for estimating the relatedness between individuals, thus opening the door to many applications of QG to decipher the architecture of quantitative traits and their response to selection; questions often off-limits in many natural systems. Genomic QG has the potential to improve our ability to mitigate species' responses to environmental change [e.g. 1], and contribute to the field of conservation genetics, such as identifying 'pre-adapted' individuals in programs of assisted migration to 'help' species keep up with climate change by estimating genetic merit of translocated individuals [e.g. 14].

From pedigrees to genome sharing

Relatedness is traditionally based on pedigrees at the base of which individuals are assumed to be unrelated (founders) [15]. Given a pedigree, it can be determined whether two alleles at a particular locus that are identical (by state or IBS) are also identical by descent (IBD) because they were inherited from the same ancestor. Due to Mendelian segregation, there is inherent variation in IBD across loci. For example, full-sibs carry, on average, 50% of their alleles IBD, but the exact percentage of genome sharing and hence their realized relatedness is subject to variation (see Fig.1A). The amount of variation around the expectations depends upon the pedigree-relationship class, and is zero only for parent and their offspring and mono-zygotic twins. Pedigree-based relatedness has well known shortcomings, such as its arbitrariness with respect to a founder population and its imprecision regarding IBD allele sharing [e.g. 16]. Nevertheless, it has been standard in the vast majority of past QG studies.

Even though different relatedness estimators that approximate IBD probabilities have been suggested in the microsatellite era, they had too large sampling errors [e.g. 17], thus it was widely

accepted that markers should only be used to check, correct, and complete pedigrees rather than replacing them [18]. With the advent of genomic data (Box 1), it has become possible to estimate the realized proportion of the genome that two individuals share IBD. The proportion of genome shared between all pairs of individuals in a population can be summarized with the genomic relationship matrix (GRM) (see Box 1 for methods).

The number of markers necessary to accurately estimate the GRM ultimately depends on the effective number of independently segregating chromosome segments that depend on the effective population size, the recombination rate and the genome size [19] and are also influenced by the mating system and population structure and history, which concertedly determine linkage disequilibrium decay. Generally speaking, species with small genomes need fewer markers than species with big genomes, where the difference between expected and realized relatedness can be high [19, 20]. A greater number of markers is also expected to be necessary in outcrossing populations in comparison to selfing, partially selfing or inbred populations because outcrossing populations generally have large effective population sizes and high recombination rates [21, 20]. It is difficult to provide a quantitative guidance for the required number of markers for a specific study. Thus, performing simulations using characteristics of the studied species and populations, and exploring multiple GRM estimation methods (see Box 1) might be the best approach.

The ‘dawn' of genomic quantitative genetics

For humans and species of economic importance, genomic data became available much earlier than for other species, thus the first genomic QG tools appeared in these fields. Many of these methods have not reached evolutionary and global change biology yet. First, we review these tools, then discuss the initial applications in natural populations.

Animal and plant breeding

Animal breeders traditionally estimated breeding values based on phenotypic scores, such as milk yield, of a focal individual and its relatives, using an animal model (Box 2). With high density SNP chips (Table 1) becoming available for many domesticated species, breeders were the first to suggest the use of the GRM in an animal model to predict genomic breeding values (GEBVs) [22] and also developed other approaches for genomic selection (GS) that even bypass the animal model (see Box 2 for details). GS revolutionized animal breeding and the field has experienced a surge of method development [23]. It was also demonstrated that an animal model coupled with a GRM instead of a pedigree leads more accurate estimates of the variance components than pedigree-based equivalents [24, 23]. Further, most pre-genomic breeding applications ignored estimating non-additive genetic effects because they generally require complex breeding designs or pedigree structures, and also because it was believed that their effects are negligible [e.g. 5]. A handful of recent empirical studies from tree breeders demonstrated that, in comparison to classical pedigree based approaches, the use of a GRM in an animal model can yield substantially more accurate separation of the additive and non-additive components of genetic variance [25, 26].

Human genetics

Human populations cannot be manipulated experimentally, so different 'non-invasive' genomic QG methods have been developed in this field. Most notably, it was suggested to exploit variation in realized genome sharing within a pedigree relationship class to estimate the heritability: Visscher *et al*. [27] proposed this idea within full-sib families, and Yang *et al*. [28] among unrelated individuals. Most importantly, these methods led to more accurate estimates of the variance components. Nevertheless, population parameters such as linkage disequilibrium between marker and quantitative trait loci can lead to different performance of this method in different species or populations. To explore this further, Visscher and Goddard [29] investigated the sampling error in the heritability estimates as a function of the sampling scheme, the sample size and the effective population size, and concluded that overall, more individuals are needed to achieve the same

precision with random population samples in comparison to pedigrees, and higher sample size is required when there is little variance in relatedness such with large effective population size or in expanding populations.

The first 'wild' applications

Applications of the genomic QG approach in wild populations have been lagging behind breeding and human genetics for obvious genomic resource reasons, even though the idea to use molecular markers to estimate heritability in natural populations had been proposed back in 1996 by Ritland [31]. Most early applications, based on a few microsatellites, yielded estimates too imprecise and often out of range [30, 31]. Recently, with dense marker panels becoming accessible in some wild pedigreed populations, it has been shown that marker-based relatedness estimates agreed with those from pedigrees (see Fig.1A and B) and also that heritability estimates based on high-density markers and pedigrees agree (Soay sheep (*Ovis aries*) [32], great tits (*Parus major)* [10]). Some empirical studies also investigated how many markers are required to estimate the heritability by comparing estimates from a decreasing proportion of the available loci. For example, in the selfing plant species, *Medicago truncatula*, heritability estimates with 25k SNPs (i.e. one marker every 10 Kb, which approximates the distance of linkage disequilibrium decay) were nearly unbiased in comparison to those obtained with >5 million SNPs (Fig.1C). Similarly, for a small island population of a wild ungulate, Soay sheep, heritability with 10k SNPs yielded unbiased estimates of the heritability obtained with 33k SNPs across several traits (Fig.1D).

The genomic QG toolbox for ecology and evolution

Bringing QG to the wild by using the pedigree-based animal model has greatly advanced our understanding of evolution [7]. Today, the acquisition of high-density marker data is possible in virtually any species (Table 1), which opens up the possibility to apply the QG research program in

the absence of a pedigree or controlled cross in any population of any species. Genomic QG also offers a greater methodological flexibility than classical QG and many evolutionary and global change research questions could benefit from it.

Genomic estimated breeding values

GEBVs could have many potential uses outside animal and plant breeding, such as in the context of conservation biology to identify suitable individuals for assisted migration or release programs. The objective of assisted migration is to 'help' species migrating to keep up with the pace of climate change, and it is becoming common practice for forest trees [14]. It has been suggested that GEBVs can be combined with gene-environment association analysis to predict phenotypes and identify changes in performance along environmental gradients [33]; Box 3 provides a proof-of-concept example of this approach in *Arabidopsis thaliana*. Since GS is based on a statistical model built on data from a 'training population' (Box 2); in order to obtain reliable predictions of the breeding values; it seems necessary to include individuals across the full environmental range over that predictions are desired. A further potential use of GEBVs is in breeding programs of endangered species, where genomic data could help to optimize breeding of remaining individuals to minimize inbreeding or outbreeding depression [34]. Releasing individuals bred in captivity into the wild is often used to restore extinct populations or help threatened populations to survive, but programs might fail if individuals are maladapted to the new environment [e.g. 35]. Identifying potentially well-adapted individuals based on their GEBVs, as described in Box 2, could be an option, or combined with allele specific genetic rescue, e.g. when disease resistance genes are known [see examples in 36]. As pointed out above, it remains unclear whether estimating GEBVs will be generally feasible in natural populations, but in an outbreeding tree species GEBVs estimated with reasonable accuracies [37]. The fact that QG estimates based on genomic markers agreed with 'standard' QG estimates in the studies on mammals, birds and plants [10, 38, 32] also indicates that it should be possible.

Realized genome sharing to quantify fitness

The GRM is just the first step in genomic QG, but it can also be used in another innovative way in evolutionary biology: The average relatedness of an individual with the population, estimated from a GRM, could be used a measure of inclusive fitness integrated across the current population as individuals whose ancestors produced many offspring should have a higher than average relatedness with the population. Such a fitness measure could also be interesting for behavioral ecology, where estimating relatedness among interacting individuals is necessary to understand social behaviors. Here, the average genomic relatedness of an individual could better estimate its inclusive fitness. Particularly, relatedness estimates based on coalescence could be adapted here [see also 39 and Box 1]. Estimating reproductive success is a challenging endeavor in wild populations for several reasons [40] but could potentially be tackled by estimating GRMs.

Evolutionary potential in the wild

Evolutionary potential is a scaled measure of the additive genetic variance (V_A); it is the heritability when it is scaled with the total phenotypic variance and evolvability when it is scaled by the trait mean [41]. Bringing genomic QG to the wild will allow assessing the evolutionary potential of a population *in situ*, which is a considerable advancement because it estimates the additive genetic variance and covariance of phenotypic traits under the conditions in which natural selection is acting, which is not the case when using common garden experiments. Such common garden estimates are not guaranteed to apply under natural conditions as V_A might vary across environments and strongly depends on population size and inbreeding levels [reviewed in 42]. Estimates of trait heritability in natural populations, however, are also prone to biases due to plastic responses to environmental variables [see 43 for a recent review]. Consequently, a good understanding of environmental but also individual-level variables, such age or body condition, that

affect the trait is necessary to account for this 'environmental' variation by including the relevant variables in the model. Alternatively, if the exact variables responsible for plastic trait variation are unknown or difficult to estimate but are known to vary spatially or temporally, appropriately modelling this spatial or temporal variation is a suitable approach. For example, in the case of an environmental variable that varies in discrete 'units', such as the effect of spring climate on timing of flowering, breeding or migration, fitting the temporal or spatial 'unit' as a (fixed or random) factor would be appropriate. If variation is more gradual, fitting an appropriate autocorrelation in the model can remove this plastic variation in traits [e.g. 44]. Another approach to remove confounding effects between genetic and environmental similarities is to estimate QG parameters from variation in relatedness within families, such as full-sibs [27], which could be possible in clutches of birds, fish or insects or half-sibs in plants [e.g. 45]. These *in situ* estimates of the evolutionary potential can be useful in many fields of ecology, evolution and conservation biology. Climate change induced new selective pressures such as increased drought, heat or seasonal shifts, which have been reported in many ecosystems but for most species we lack *in situ* or even any estimates of V_A at ecologically key traits including phenology or drought tolerance [e.g. 46, 47]. Climate change is also expected to drive changes in community assemblages due to different dispersal abilities at different species, as has been shown for alpine plants [48]. *In situ* evaluation of the evolutionary potential of key life-history traits is necessary to assess whether these species will be able to adapt to the changed competition regime.

Estimates of V_A could also be particularly useful for models of eco-evolutionary dynamics (Hendry 2016). For example, Ellner [49] suggested a framework to partition change in a population into contributions of evolution, non-heritable trait change and environmental. To date, this theoretical framework has been limited to experimental systems, but with genomic QG it could be applied in natural settings.

With the decreasing costs of genotyping, measuring phenotypes of individuals can become the limiting factor for QG studies in natural populations. Although the variance in (genomic) breeding

values is not V_A [see e.g. 50], it could be used as a proxy for it, which would allow prediction of evolutionary potential in new populations without the need to phenotype individuals. However, genotype-by-environment interactions might limit the application of this approach to species where reaction norms are known across environmental gradients.

Predicting response to selection

The evolutionary potential informs about the capacity of a population to respond to selection. Knowledge of V_A for a trait or for fitness directly applies to predictive QG models of eco-evolutionary dynamics. This is especially relevant in the context of predicting shifts of species' ranges caused by climate change, and many voices have called for the development of eco-evolutionary forecasting models in that context [e.g. 51]. Quantitative genetics theory provides a well formed framework for predicting eco-evolutionary dynamics in single or among competitively interacting species [e.g. 52]. Current research efforts provide practical examples of implementation of this framework, together with modeling of species ecological niche distributions, to predict changes of species' distribution in Drosophilid flies in Australia [53] or endemic alpine plants in Austria [54]. However, this approach is data hungry and needs estimates of V_A for climate-adaptation traits and the selection gradients acting on them. There is hope that by using genomic QG tools such as environment-related GEBVs in natural populations (Box 3) we will be able to alleviate the dearth of estimates of QG parameters necessary to correctly forecast the fate of natural species facing a warming climate.

Conclusions

Genomic QG has the potential to advance our understanding of evolutionary processes in natural populations in several ways. Its main advantage is that we no longer need pedigrees or breeding experiments for QG analysis, which means that we can study additive genetic (co)variances, but

also selection, in virtually any species in a natural setting. Genomic QG parameters estimates can be more precise than pedigree-based estimates, because they have fewer assumptions and offer a greater methodological flexibility. This ability to conduct quantitative genetic analyses in natural populations could allow more accurate predictions of species' persistence under climate change and enable informed conservation strategies. As we pointed out, the feasibility of this approach will partly depend on the specific study system. The relative ease with which high-density markers can be obtained make us cautiously optimistic about the potential of marker-based quantitative genetics to considerably contribute to our understanding of evolutionary processes in natural populations and mitigating climate change.

Acknowledgements

We thank Thomas Mitchell-Olds, Jean-Luc Jannink, Marcel Visser and two anonymous reviewers for constructive comments on the manuscript. The idea for this manuscript developed at Monte Verità Conference in 2016 on "The genomic basis of eco-evolutionary change" organized by the ETH Zurich's center for Adaptation to a changing environment (ACE). We would like to acknowledge the organizers for providing such a fruitful, scientific atmosphere. Camillo Bérénos, Josephine Pemberton, John Stanton-Geddes and Peter Tiffin kindly allowed us to re-use their figures and provided the data to do so. SF was partly supported by the Swiss National Science Foundation (project 31003A_160123). FG is supported by grant PP00P3_144846 from the Swiss National Science Foundation. KC was supported by an ACE Fellowship while working on the manuscript.

Reference list

1 Gonzalez, A., *et al.* (2013) Evolutionary rescue: An emerging focus at the intersection between ecology and evolution. *Phil. Trans. R. Soc. B* 268, 20120404

2 Fisher, R.A. (1918) The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society* 52, 399-433

3 Falconer, D.S. and Mackay, T.F. (1996) *Introduction to quantitative genetics*. Longman

4 Nelson, R.M., *et al.* (2013) A century after fisher: Time for a new paradigm in quantitative genetics. *Trends in Genetics* 29, 669-676

5 Hill, W.G. (2010) Understanding and using quantitative genetic variation. *Phil. Trans. R. Soc. B* 365, 73-85

6 Henderson, C.R. (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, 423-447

7 Charmantier, A., *et al.*, eds (2014) *Quantitative genetics in the wild*. Oxford University Press

8 Visscher, P.M., *et al.* (2017) 10 years of gwas discovery: Biology, function, and translation. *The American Journal of Human Genetics* 101, 2-22

9 Gienapp, P., *et al.* (2017) Environment-dependent genotype-phenotype associations in avian breeding time. *Frontiers in Genetics* 8

10 Robinson, M.R., *et al.* (2013) Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. *Mol. Ecol.* 22, 3963-3980

11 Santure, A.W., *et al.* (2015) Replicated analysis of the genetic architecture of quantitative traits in two wild great tit populations. *Mol. Ecol.* 24, 6148-6162

12 Hill, W.G. (2012) Quantitative genetics in the genomics era. *Curr. Genom.* 13, 196-206

13 Travisano, M. and Shaw, R.G. (2013) Lost in the map. *Evolution* 67, 305-314

14 Aitken, S.N. and Whitlock, M.C. (2013) Assisted gene flow to facilitate local adaptation to climate change. *Ann. Rev. Ecol. Evol. Syst.* 44, 367-388

15 Lynch, M. and Walsh, B. (1998) *Genetics and analysis of quantitative traits*. Sinauer

16 Speed, D., *et al.* (2012) Improved heritability estimation from genome-wide snps. *Am. J. Hum. Genet.* 91, 1011-1021

17 Csilléry, K., *et al.* (2006) Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics* 173, 2091-2101

18 Pemberton, J.M. (2008) Wild pedigrees: The way forward. *Proceedings of the Royal Society B-Biological Sciences* 275, 613-621

19 Hill, W.G. and Weir, B.S. (2011) Variation in actual relationship as a consequence of mendelian sampling and linkage. *Genet. Res.* 93, 47-64

20 Wang, J. (2016) Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient? *Theoretical Population Biology* 107, 4-13

21 Gay, L., *et al.* (2013) Pedigree-free estimates of heritability in the wild: Promising prospects for selfing populations. *PLoS ONE* 8, e66983

22 Meuwissen, T., *et al.* (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819-1829

23 Meuwissen, T., *et al.* (2016) Genomic selection: A paradigm shift in animal breeding. *Anim. Front.* 6, 6-14

24 Jannink, J.-L., *et al.* (2010) Genomic selection in plant breeding: From theory to practice. *Brief. Funct. Genom.* 9, 166-177

25 Muñoz, P., *et al.* (2014) Unraveling additive from nonadditive effects using genomic relationship matrices. *Genetics* 198, 1759-1768

26 Gamal El-Dien, O., *et al.* (2016) Implementation of the realized genomic relationship matrix to open-pollinated white spruce family testing for disentangling additive from nonadditive genetic effects. *G3* 6, 743-753

27 Visscher, P.M., *et al.* (2006) Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* 2, e41

28 Yang, J., *et al.* (2010) Common snps explain a large proportion of the heritability for human height. *Nature Genet.* 42, 565-569

29 Visscher, P.M. and Goddard, M.E. (2015) A general unified framework to assess the sampling variance of heritability estimates using pedigree or marker-based relationships. *Genetics* 199, 223-232

30 Coltman, D.W. (2005) Testing marker-based estimates of heritability in the wild. *Mol. Ecol.* 14, 2593-2599

31 Garant, D. and Kruuk, L.E.B. (2005) How to use molecular marker data to measure evolutionary parameters in wild populations. *Mol. Ecol.* 14, 1843-1859

32 Bérénos, C., *et al.* (2014) Estimating quantitative genetic parameters in wild populations: A comparison of pedigree and genomic approaches. *Mol. Ecol.* 23, 3434-3451

33 Lasky, J.R., *et al.* (2015) Genome-environment associations in sorghum landraces predict adaptive traits. *Science Advances* 1, e1400218

34 Frankham, R. (2008) Genetic adaptation to captivity in species conservation programs. *Mol. Ecol.* 17, 325-333

35 Griffiths, R.A. and Pavajeau, L. (2008) Captive breeding, reintroduction, and the conservation of amphibians. *Cons. Biol.* 22, 852-861

36 Harrisson, K.A., *et al.* (2014) Using genomics to characterize evolutionary potential for conservation of wild populations. *Evol. Appl.* 7, 1008-1025

37 Beaulieu, J., *et al.* (2014) Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity* 113, 343-352

38 Stanton-Geddes, J., *et al.* (2013) Estimating heritability using genomic data. *Meth. Ecol. Evol.* 4, 1151-1158

39 Rousset, F. (2002) Inbreeding and relatedness coefficients: What do they measure? . *Heredity* 88, 371-380

40 Shaw, R.G., *et al.* (2008) Unifying life-history analyses for inference of fitness and population growth. *Am. Nat.* 172, E35-E47

41 Houle, D. (1992) Comparing evolvability and variability of quantitative traits. *Genetics* 130, 195-204

42 Hoffmann, A.A., *et al.* (2017) Revisiting adaptive potential, population size, and conservation. *Trends Ecol. Evol.* Online Early

43 Wood, C.W. and Brodie III, E.D. (2016) Evolutionary response when selection and genetic variation covary across environments. *Ecol. Lett.* 19, 1189-1200

44 Germain, R.R., *et al.* (2016) Direct and indirect genetic and fine-scale location effects on breeding date in song sparrows. *J. Anim. Ecol.* 85, 1613-1624

45 Bontemps, A., *et al.* (2016) In situ marker-based assessment of leaf trait evolutionary potential in a marginal european beech population. *J. Evol. Biol.* 29, 514-527

46 Alberto, F.J., *et al.* (2013) Potential for evolutionary responses to climate change – evidence from tree populations. *Glob. Change Biol.* 19, 1645-1661

47 Gienapp, P., *et al.* (2014) Why climate change will invariably lead to selection on phenology. *Proceedings of the Royal Society B-Biological Sciences* 281, 20141611

48 Alexander, J.M., *et al.* (2015) Novel competitors shape species' responses to climate change. *Nature* 525, 515-518

49 Ellner, S.P., *et al.* (2011) Does rapid evolution matter? Measuring the rate of contemporary evolution and its impacts on ecological dynamics. *Ecol. Lett.* 14, 603-614

50 Hadfield, J.D., *et al.* (2010) The misuse of blup in ecology and evolution. *Am. Nat.* 175, 116-125

51 Urban, M.C., *et al.* (2016) Improving the forecast for biodiversity under climate change. *Science* 353, aad8466

52 Duputié, A., *et al.* (2012) How do genetic correlations affect species range shifts in a changing environment? *Ecol. Lett.* 15, 251-259

53 Bush, A., *et al.* (2016) Incorporating evolutionary adaptation in species distribution modelling reduces projected vulnerability to climate change. *Ecol. Lett.* 19, 1468-1478

54 Cotto, O., *et al.* (2017) A dynamic eco-evolutionary model predicts slow response of alpine plants to climate warming. *Nature Comm.* In press.

55 Pérez-Enciso, M. (2014) Genomic relationships computed from either next-generation sequence or array snp data. *J. Anim. Breed. Genet.* 131, 85-96

56 Dodds, K.G., *et al.* (2015) Construction of relatedness matrices using genotyping-by-sequencing data. *BMC Genom.* 16, 1047

57 Palaiokostas, C., *et al.* (2016) Genomic prediction of resistance to pasteurellosis in gilthead sea bream (*sparus aurata*) using 2b-rad sequencing. G3 (bethesda) 1–8 (2016). *G3* 6, 3693-3700

58 Morrison, A.C., *et al.* (2013) Whole-genome sequence-based analysis of high-density lipoprotein cholesterol. *Nature Genet.* 45, 899-901

59 Snyder-Mackler, N., *et al.* (2016) Efficient genome-wide sequencing and low-coverage pedigree analysis from noninvasively collected samples. Early online, DOI: 10.1534/genetics.1116.187492

60 Ackerman, M.S., *et al.* (2017) Estimating seven coefficients of pairwise relatedness using population genomic data. *Genetics* 206, 105-118

61 Meynert, A.M., *et al.* (2014) Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinform.* 15, 247

62 Zhuang, Z., *et al.* (2012) Detecting identity by descent and homozygosity mapping in whole-exome sequencing data. *PLoS ONE* 7, e47618

63 Milligan, B.G. (2003) Maximum-likelihood estimation of relatedness. *Genetics* 163, 1153-1167

64 Goddard, M.E., *et al.* (2011) Using the genomic relationship matrix to predict the accuracy of genomic selection. *J. Anim. Breed. Genet.* 128, 409-421

65 Powell, J.E., *et al.* (2010) Reconciling the analysis of ibd and ibs in complex traits. *Nature Rev. Genet.* 11, 800-805

66 Lee, S.H., *et al.* (2013) Estimation of snp heritability from dense genotype data. *The American Journal of Human Genetics* 93, 1151-1155

67 Browning, B.L. and Browning, S.R. (2013) Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* 194, 459-471

68 Purcell, S., *et al.* (2007) Plink: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559-575

69 Yang, J., *et al.* (2011) Gcta: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76-82

70 Anastasio, A.E., *et al.* (2011) Source verification of mis-identified arabidopsis thaliana accessions. *Plant J.* 67, 554-566

71 Fournier-Level, A., *et al.* (2011) A map of local adaptation in arabidopsis thaliana. *Science* 334, 86-89

72 Horton, M.W., *et al.* (2012) Genome-wide patterns of genetic variation in worldwide arabidopsis thaliana accessions from the regmap panel. *Nature Genet.* 44, 212-216

73 Hijmans, R.J., *et al.* (2005) Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965-1978

74 Endelman, J.B. (2011) Ridge regression and other kernels for genomic selection with r package rrblup. *Plant Genome* 4, 250-255

448 Table 1 Overview over high-throughput genotyping approaches and sequencing approaches that are potentially suitable to estimate relatedness.

Data type	Genomic resources needed	Pros	Cons	Cost	Examples of studies applying this approach to calculate the GRM
SNP chip	Reference genome	"Ready to use" - No bioinformatics work	Availability mostly limited to model and breeding systems, and possibly close relatives. Ascertainment bias: new populations can have different polymorphisms [55]	Low, if chip available	animal and plant breeding [24, 23], human genetics [28], wild species [32]
RRL- or RAD-sequencing	None, but improved if there is a reference genome	Flexibility of experimental design assisted by software predictions on number of markers. Established bioinformatics for data	Bias on relatedness estimates due to allele dropout and high proportion of missing data [56]	Medium	[56, 57]

		analysis with easy-to-use pipelines			
Whole-genome sequencing	Reference genome	High-confidence genome-wide information and use of established bioinformatics tools	High costs for large genomes, and can yield excessive amount of information [38]	Very high	[58, 38]
Low-coverage whole-genome sequencing	Reference genome	Dense genome representation including causal variants that can be suitable for parallel association studies.	Requires bioinformatics knowledge to use software building on probabilistic framework.	High	[59, 60]
Sequence or Exome capture	Reference genome of closely related species or	Targeted sequencing of selected loci, which can include causal variants	Requires high-quality genomic resources for bait design and to minimize allele bias [61]	High	[62]

	transcriptome	suitable for parallel association studies.		
Genome skimming	Reference genome	Minimal sequencing effort to survey high-copy regions of the genome, typically organellar and ribosomal DNA	No studies yet built a GRM. Complex evolution of high-copy nuclear loci can limit applicability in natural populations.	Low

BOXES

BOX 1 Methods to estimate the genomic relationship matrix (GRM)

1. Approximate identity-by-descent (IBD) probabilities at single locus. This approach includes a whole class of maximum-likelihood [e.g. 63, 60] and method of moments estimators [see 17 for an overview].

2. Cumulative excess of recent coalescence. A coalescent tree is used to trace the origin of alleles in the current population back to a most recent common ancestor (MRCA). Rousset [39], suggested to define kinship coefficients as the degree of excess coalescent time to their MRCA at each independently segregating segments of the genome, however, no estimator or software have been proposed yet.

3. Average allelic correlation (a.k.a. unified additive relationship). This is the most standard class of approaches to estimate the GRM in animal and plant breeding and human genetics (see Table I for software). In breeding, the method appeared with the uprise of genomic selection (GS), and calculated as the correlation between genotypes of two individuals across all loci [64]. In human genetics, the approach appeared in Yang et al. [28], and was explained in Powell et al. [65]. The approach circumvents the problem of the arbitrarily reference population by defining it as the current population. As a result, IBD is no longer a probability, but a correlation coefficient that can be used to predict if a gamete carries the same allele as another irrespective of its origin (i.e. IBS). Adjustment for linkage disequilibrium (LD) been proposed by [16], which can lead to improved estimates , especially with sparse SNP data [66]. When using dense SNP data, LD-adjustment attributes too much weight to the SNPs with low minor allele frequencies.

4. *Haplotype sharing.* Haplotype sharing approaches require both marker data and a genetic map [e.g. 67]. A common approach in GS is to use dense SNP chip on a few key individuals to identify haplotypes, and a sparse chip on the rest of the individuals [see review by 23].

474 Table I. Examples of software to estimate the GRM

Software	PLINK (v1.07)	GCTA	LDAK	BEAGLE
Method	1	3	3	4
Reference (software)	[68]	[69]	[16]	[67]
Website	http://zzz.bwh.harvard.edu/plink/	http://cnsgenomics.com/software/gcta/	http://dougspeed.com/ldak/	http://faculty.washington.edu/browning/beagle/beagle.html

475

476

BOX 2 Two key statistical models to estimate QG parameters using genomic data

The two main types of statistical models dominate the QG literature and both can make use of genomic data. In the animal model, the pedigree can be replaced with the GRM, and in genome regressions, genomic data is a necessary ingredient.

	Animal model	Genome regression
Key reference	Henderson (1975)	Meuwissen, Hayes, Goddard (2001)
The model	$y_i = \mu + a_i + e_i$ where a_i is the breeding value of individual i. Since a_i's are unknown, they are estimated from the covariance between relatives in additive genetic effects given by 2*kinship*V_A, thus using a pedigree or a GRM	$y_i = \mu + \sum(SNP_{i,j} * g_j) + e_i$ where $SNP_{i,j}$ is the genotype of individual i at loci j, and g_j is the effect of the SNP on the phenotype
explanation of common model terms	y_i is a trait value of individual i, μ is the population mean, and e_i is the residual term	
Breeding value (BV) estimation	Using the best linear unbiased predictor (BLUP) in a mixed effects model	BLUPs assuming equal variances across SNPs is possible, but Bayesian methods generally work better

About the animal model

Animal model is most often used to estimate variance components, such as the additive genetic variance, which serves to estimate the heritability of traits or genetic correlations between traits. Breeding values are the estimated random effects and can be used to characterize individuals, however, several authors warn not to use them in further statistical analysis without incorporating their estimated standard errors [e.g. 50], e.g. in a hierachical Bayesian framework. Generally large pedigrees are required but it can also be used with a common garden experimental data.

About genome regressions

Genome regressions are principally used in a prediction context to help breeders identify superior individuals. Thus, the model is fitted in a training population to estimate breeding values of individuals of unknown trait value in a candidate population (such as phenotyping milk yield at an early age or at a male cow). GEBVs are generally much more accurate than EBVs. Particularly high accuracy is expected with a large number of individuals in the training population, for a highly heritable trait, and when the number of segregating chromosome segments is small (i.e. small genome and small effective population size). GS does not work well across breeds. Instead, the best performance is achieved when individuals in the candidate population are related to individuals in the training population.

Box 3 Using genomic relatedness and home environments to identify locally adapted genotypes

A common challenge in conservation genetics is to identify locally-adapted genotypes for a given site. Here, we demonstrate that GRMs for wild accessions collected across diverse climates ('ecotypes') can be applied to predict which genotypes are adapted to cold sites and which are adapted to warm sites. Our approach relies on a three-way correlation between genomic similarity (GRMs), similarity in home environments (winter cold), and similarity in locally adaptive traits (unobserved and predicted). As such, correlation between population structure and environmental gradients causing local adaptation is not a problem for our approach, rather this correlation is leveraged to predict adaptive trait variation. By contrast, population structure-environment correlation is problematic for genome-environment association studies that seek to identify individual causal loci involved in local adaptation. We reanalyzed published data on *Arabidopsis thaliana* [70-72] and global climate [73]. We used >200k SNPs [72] to estimate a GRM based on average allelic correlation (Box 1). We fitted an animal model where an ecotype's native winter cold environment (minimum winter temperature) was used as 'phenotype' and modeled as a function of the GRM with 10-fold cross validation to predict native temperatures as GEBVs [74]. Based on the assumption that populations are locally adapted along temperature gradients, these GEBVs can also be interpreted as predictions of a 'latent cold-adapted phenotype', *i.e.* an unobserved phenotype assumed to confer high fitness in particular temperatures. We next show that these predictions closely correspond to which genotypes are locally adapted along temperature gradients. Although here we validated our predictions with experimental data, note that no experiments were used to generate our predictions; predictions are based solely on the relationship between GRM and native environments. Individual relative fitness of 157 ecotypes was measured in four common gardens by Fournier-Level et al. [71]. In each garden, we regressed relative fitness against predicted latent cold-adapted phenotypes to identify the expected most fit latent phenotype at each site (Fig. I, data from two sites shown in insets, curve shows fitted quadratic regression).

The fittest latent cold-adapted phenotype at each site varied closely with differences in winter cold among sites (central panel, $R^2 = 0.92$). This relationship indicates that the GRM association with native environments can be used to predict which genotypes are adapted to cold sites and which are adapted to warm sites, information that might guide conservation and restoration actions.

Glossary

Infinitesimal model: The infinitesimal model assumes that the genetic component of a trait is determined by an infinite number of unlinked genes that each has an infinitesimally small, additive effect. This will lead to continuous genetic variation in the trait and this model is hence generally used to model so-called 'quantitative' or 'complex' traits. It is, however, also applicable to discrete traits, as e.g. litter size, and even binary traits if trait expression depends on one or more thresholds that the summed allelic effects need to pass.

Heritability: The narrow-sense heritability (h^2) is the ratio of the additive genetic variance (V_A) to the total phenotypic variance (V_P) of a trait: $h^2 = \frac{V_A}{V_P}$. The narrow-sense heritability is used to predict the response to selection of a trait when using the breeders' equation. When only the whole genetic variance (V_G) is known, which also includes non-additive genetic effects (dominance and epistasis), broad-sense heritability (H^2) is estimated.

Evolvability can be defined in the broader and narrower sense. In the broader sense it simply is the ability of a trait to evolve (including from novel mutations). In the narrower sense it is the additive genetic variance scaled by the trait mean: $volvability = \frac{V_A}{trait\ mean^2}$, or the additive genetic coefficient of variation (CV_A): $CV_A = \frac{V_A}{trait\ mean}$.

Quantitative trait locus: A quantitative trait locus (QTL) is the genomic section, i.e. a locus, that contribute to the variation (together with other QTL) of a quantitative trait, i.e. a trait that shows continuous phenotypic variation.

Genome-wide association study: A genome-wide association study (GWAS) aims to identify loci determining a trait by testing for associations between molecular markers and trait values. In the common case when SNP chips are used, the underlying assumption is that the markers are in tight linkage with or within the loci responsible for trait variation.

Single nucleotide polymorphism: A single nucleotide polymorphism (SNP) is a variation at a specific genomic position. For example, at a specific position in a genome, the base C appears in most individuals, but in a minority the base A appears. This specific SNP would have the two nucleotide variations – C or A – as its alleles.

Identity-by-descent: A genomic section is identical by descent (IBD) in two individuals if they have inherited it from a common ancestor.

Identity-by-state: A genomic section is identical by state (IBS) in two individuals if it has the same nucleotide sequence in these two individuals. IBS can arise by chance (mutation) or because the individuals inherited it from the same ancestor, in which case it would also be identical by descent.

Realized relatedness: The relatedness of two individuals calculated from a pedigree (kinship) is their expected genetic similarity (or proportion of genome sharing). There is, however, variation around this expectation (except for parent-offspring pairs and monozygotic twins) because of the segregation of independent genome segments during meiosis (mendelian noise). This variation is lowered by more independently segregating segments[e.g. 27, 19]. An interesting consequence of this is that for species with 'small' genomes, i.e. with fewer independently segregating segments, QG estimates based on realized relatedness will be more precise than estimates based on expected relatedness.

Genomic relationship matrix: The pedigree-based relatedness matrix, often denoted A, contains all pairwise relatedness estimates (which are two time the kinship coefficient θ) in the off-diagonals and the individuals inbreeding coefficient in the diagonal. The genomic relatedness matrix (GRM), often also denoted G, is the same but with the relatedness estimates derived from genomic markers. Various approaches to estimate relatedness from markers exists (see Box 1) and GRMs can be constructed with any of them.

Linkage disequilibrium is the non-random association of alleles at different loci. Loci are in linkage disequilibrium, or linked, if certain combinations of alleles occur more often than expected

by chance. Linkage disequilibrium can be caused by a multitude of factors, including selection, recombination rate, mating system, population structure or physical proximity in the genome.

Non-additive effects include dominance and epistasis. Dominance means that the effect of one allele masks (is 'dominant' over) the effect of the other (recessive) allele. Incomplete and co-dominance are cases when the dominant allele is not completely masking the recessive allele and the heterozygous phenotype is somewhat intermediate between the homozygous dominant and homozygous recessive phenotype. Epistasis, or gene-gene interactions, occurs when the allelic effect at one locus depend on the alleles of one or more other loci, the genetic background.

Assisted migration: Many species have limited dispersal abilities or their dispersal is impaired by habitat fragmentation, which means that they cannot reach other, isolated populations or currently unoccupied but suitable habitat. In assisted migration, or assisted gene flow, individuals are artificially moved from one population to improve another population's probability to persist or to colonize novel but suitable habitat, e.g., in the context of climate change.

Figure caption

Fig.1 Genomic relatedness and heritability estimates. (A) Relationship between marker-based and pedigree relatedness in one of the wild study populations with the deepest pedigree, the St. Kilda soay sheep population. Realized or 'genomic' relatedness, estimated from ca. 37000 SNP markers, is plotted against IBD relatedness, estimated from the pedigree. The variation of realized relatedness estimates with the same IBD relatedness class, seen as vertical spread, is due to Mendelian segregation (and sampling error). (B) shows, for the same population, that the correlation between pedigree and marker-based relatedness estimates increases with the number of SNPs used to estimate the GRM. (C) and (D) show the effect of number of markers on heritability estimates in *Medicago truncatula* and soay sheep, respectively. Boxes and whiskers indicate the variation in estimates from the repeated samples of the reduced data sets. When few markers are used to estimate the GRM heritability, estimates are down-ward biased. Increasing the number of markers reduces this bias, but, at some point, adding more markers has little effect on the estimate anymore, only on its variation. In (B) and (D) the x-axis shows the proportion of SNPs used in relation to the complete data set and the absolute numbers are 926, 1852, 3704, 11111, 18518, 25926, respectively. (A), (B) and (D) re-drawn after [32]; (C) re-drawn after [38].