

Perspective: conservation genetics enters the genomics era

John C. Avise

Received: 8 September 2009 / Accepted: 12 October 2009 / Published online: 15 November 2009
© Springer Science+Business Media B.V. 2009

Introduction

Throughout most of my professional career (which began in the late 1960s), a long-sought Holy Grail in molecular ecology and evolution was to obtain extensive nucleic acid sequences from large numbers of loci and organisms. In lieu of efficient DNA-sequencing technologies, researchers adopted a succession of less direct approaches for estimating various genomic parameters such as heterozygosities, kinship coefficients, or genetic distances. These laboratory techniques included allozyme electrophoresis (mid-1960s), the immunological approach of micro-complement fixation (1960s), gel-sieving and other methods to reveal hidden protein variation (early 1970s), restriction-enzyme assays especially of mitochondrial DNA (late 1970s), DNA/DNA hybridization (1970s), DNA fingerprinting by minisatellites (1980s), PCR-based sequencing of particular target genes for which conservative primers were developed (late 1980s), RAPD (randomly amplified polymorphic DNA) assays (1990s), microsatellite analyses (1990s), DNA barcoding based on a mitochondrial gene (2000s), and several other molecular approaches for revealing genetic variation in particular proteins or classes of nucleic acids (see Hillis et al. 1996; Avise 2004; Freeland 2005).

Different laboratory methods yielded genetic markers well-suited for addressing different sections along a phylogenetic spectrum from the micro- to the macro-evolutionary: detection of clonal identity or non-identity (e.g., via DNA fingerprinting or multi-locus allozymes), population demography and mating systems (allozymes,

microsatellites), intraspecific population structure and phylogeography (allozymes, mtDNA), speciation processes and species differences (barcoding, allozymes, mtDNA), hybridization and introgression (allozymes, mtDNA), and supra-specific phylogenetics at many temporal scales (via microcomplement fixation, DNA-DNA hybridization, and DNA sequencing of particular nuclear or cytoplasmic loci). In many cases, the data also served to improve our mechanistic understanding of a wide range of molecular-level phenomena such as mutation rates and patterns, gene duplications, the phenomenon of concerted evolution, and the operation of natural selection on particular loci. The primary limitation of most methods (with the possible exception of DNA hybridization) was that only a tiny fraction of the genome was accessible from which to make estimates of the genome-wide parameters that ultimately were of interest.

In recent years, later-generation molecular technologies have made mass-scale nucleic acid sequencing almost routine. For example, by the spring of 2009, at least one entire genome had been sequenced from each of about 1,000 species (including 100 eukaryotes), with another 1,000 species in various stages of sequence completion. Modern molecular methods such as 454 pyrosequencing also make it possible to sequence thousands of protein-coding genes using expressed sequence tags (ESTs) from the transcriptomes (messenger RNA pools) of multiple individuals, even in non-model organisms (Papanicolaou et al. 2005; Hudson 2007). Furthermore, in this “genomics revolution,” dramatic advancements in microchip arrays and related technologies have made gene-expression profiling (transcriptomics, proteomics, and metabolomics) practicable at unprecedented genomic scales (Gibson and Muse 2009). Indeed, molecular technologies are no longer *the* limiting factor in genetic analysis, often having been

J. C. Avise (✉)
Department of Ecology and Evolutionary Biology,
University of California, Irvine, Irvine, CA 92697, USA
e-mail: javise@uci.edu

replaced by issues related to each researcher's time, energy, and capacity to synthesize and interpret vast quantities of genomic data. Conservation genetics has long maintained close collaborative contact with molecular biology (Schonewald-Cox et al. 1983; Avise and Hamrick 1996; Smith and Wayne 1996). Here I will briefly speculate on how the field of conservation genetics might be impacted by the genomics revolution. Some of my thoughts in the sections that follow were motivated by talks and posters at an international symposium (*Integrating Population Genetics and Conservation Biology*), organized by the ESF Networking Programme CONGEN and held in Trondheim, Norway, May 23–26, 2009 (see Ouborg 2009).

Background

In conservation genetics, molecular data can play two fundamental roles that I will refer to as the mechanistic (or functional) and the inventorial. With respect to the mechanistic role, the genomics revolution will open countless opportunities to improve our understanding of genetic and cellular operations and their ramifications for organismal development, ecology, and evolution. With respect to the inventorial role, the genomics revolution will vastly improve our capacity to take genealogical stock of biological resources at all levels in the phylogenetic hierarchy, ranging from individuals and demes to populations, species, and higher taxa. These two basic roles are complementary, potentially synergistic, and will find many applications in conservation genetics.

A distinction between the mechanistic and inventorial roles for molecular data can be traced to the base of biology's molecular revolution in the mid-1960s. Soon after researchers introduced allozyme methods to population biology (Hubby and Lewontin 1966), a debate arose between the neutralists and the selectionists with regard to the evolutionary significance of the newly discovered genetic variation. Neutralists argued that molecular variation was mostly irrelevant to organismal fitness, whereas selectionists suspected that most molecular variation (certainly at the protein level) was visible to natural selection and thus highly germane to the adaptive process. This controversy resurfaced time and again as biologists contemplated each new type of molecular data provided by the latest laboratory method. Relevant research typically proceeded on two fronts: testing various mathematical predictions of neutrality theory against observed magnitudes or pattern of molecular heterogeneity in various species; and addressing the functional properties of particular genes and alleles more directly. The selection-neutrality controversy in molecular ecology and evolution led to what I am now categorizing as the field's longstanding mechanistic

orientation. The guiding question that motivates this research paradigm is, "What is the functional significance of molecular variation?"

An equally important inventorial role for molecular variation also emerged in the mid-1960s. Under this paradigm, appropriate genetic variation (whether strictly neutral or not) can be genealogically or phylogenetically informative in various ecological, behavioral, and evolutionary arenas. Suitable molecular variation can reveal, for example, the genetic parentage of particular offspring in the wild, or the spatial genetic structures of conspecific populations, or the phylogenetic relationships of species and higher taxa. Such applications in molecular ecology and evolution epitomize the field's longstanding inventorial orientation, which is guided by the question: "What can molecular markers unveil about organismal kinship, natural history, behavior, and phylogeny?"

Today, the distinction between the functional and inventorial roles for molecular variation continues to find expression in the differing research paradigms of different genomics laboratories. For example, in the genomics era, standard screening for thousands or even millions of SNPs (single-nucleotide polymorphisms) and other genetic variants (Kendal 2003) has become possible for many model as well as non-model species (e.g. Pertoldi et al. 2010). For researchers interested in the natural-history side of conservation genetics, this newfound wealth of molecular markers will permit refined studies of genetic parentage, geographic population structure, hybridization, introgression, and other such biological phenomena that often find conservation relevance. But for researchers focused on genetic function, these data (including linkage patterns) are greeted with excitement because they should help to clarify the ecological and evolutionary forces that shape genomic architectures. For example, the data should yield improved estimates of genic heterozygosity (H) within individuals and thereby help unveil mechanisms underlying the long-discussed relationship between genetic variation and individual genetic fitness (Mitton 1997; Frankham et al. 2002). Previous empirical attempts to estimate H values in natural populations probably came from too few genes to reliably rank-order individuals with respect to heterozygosity (Mitton and Pierce 1980), but the refined estimates from thousands of loci should permit researchers to overcome this limitation and thereby help address questions of the following sort: Do the observed fitness effects stem from heterosis at particular loci, or from genome-wide variation *per se*? Answers to this and related questions are relevant to functional workings of the genome, and also, ultimately, to conservation genetics.

A subtle tension between mechanistic and inventorial paradigms is similarly evident in other areas of conservation genetics. With respect to intraspecific variation, for

example, researchers interested in estimating effective population size (N_e) from molecular data normally prefer to monitor variation in neutral markers, whereas researchers interested in adaptive processes might show greater research interest in understanding the dynamics of loci under strong selection (such as MHC loci in mammals or self-incompatibility loci in plants). With respect to longer-term evolution, researchers focused on phylogenetic reconstruction tend to view neutral molecular markers as informative signal, and genes under intense selection as potential phylogenetic noise (homoplasy); whereas researchers with a mechanistic orientation tend to view genes under selection as being of special interest because of their relevance to adaptive evolution. Both of these world-views have merit, of course, and indeed the deepest evolutionary insights often emerge from integrating the two. For example, a powerful approach to understanding the evolution of adaptive traits is to map selected characters onto phylogenies estimated from neutral markers (Avisé 2006). Furthermore, genes potentially under strong selection are often first identified because they have particular features (such as exceptionally high or low F_{st} values, or perhaps high ratios of non-synonymous to synonymous nucleotide substitutions) that make them stand out from the crowd of otherwise neutral or nearly neutral loci.

From conservation genetics to conservation genomics

The genomics revolution will improve scientific capabilities within both the mechanistic and the inventorial traditions of conservation genetics. An excellent example combining both arenas involves ongoing research (detailed at the Trondheim symposium by Chris Wheat) on the Glanville fritillary butterfly (*Melitaea cinxia*), an organism for which extensive ecological and natural history information (but not yet a fully sequenced genome) are available (Ehrlich and Hanski 2004). The researchers used 454 pyrosequencing to study hundreds of thousands of expressed sequence tags (ESTs) in samples from a metapopulation in Finland (Vera et al. 2008), and integrated the genomic information both with field data on individual dispersal and with physiological parameters related to flight. The data are proving to be highly informative not only about the population genetic structure and metapopulation dynamics of this species, but also about genetic variation that functionally underlies individual differences in flight metabolism and dispersal capabilities.

In this special issue of *Conservation Genetics* (dedicated to the Trondheim symposium), Joop Ouborg expounds at greater length on many of the research opportunities in conservation genetics that fall within the functional or mechanistic paradigm of the genomics revolution (Ouborg

et al. 2010). So, here I will focus instead on what I perceive to be some special research opportunities on the inventorial side of conservation genomics. The first and most obvious point to be made is that additional molecular markers will mean improved estimates of various genomic parameters such as individual heterozygosities and genetic distances. From the thousands of loci made accessible for analysis by the genomics revolution, we can expect, for example, greater statistical power (i.e., higher exclusion probabilities) for assessing genetic paternity and maternity, and likewise much greater power for assessing population structure, introgression, and phylogenetic relationships among taxa. However, any unbridled enthusiasm for such gains should be tempered by the realization that traditional molecular markers already provide at least adequate power for providing genealogical inventories of many biological phenomena relevant to conservation. For example, parentage analyses via conventional microsatellite markers routinely entail exclusion probabilities >99% in suitable biological settings (such as when one parent is already known or suspected from independent evidence); and hybridization and introgression can be detected readily between many species pairs, and dissected using cytonuclear analyses (Avisé 2001) as applied to data from standard nuclear and mitochondrial markers. Thus, in such cases the benefits to be derived from the genomics revolution will often be matters of degree rather than unprecedented breakthroughs.

Nevertheless, I do see at least three broad arenas in which the genomics revolution might lead to qualitative breakthroughs within the inventorial research paradigm of conservation genetics.

Three opportunities for transformational research

The first of these arenas is in assessing relative levels of genetic kinship between pairs of individuals within local demes. Traditional genetic markers have served the field of population genetics quite well on issues of clonal identity/non-identity, genetic paternity and maternity, and broader population structures, but they have been almost useless in distinguishing, for example, first-cousins from second-cousins or other close categories of genetic kinship (where the theoretical coefficients of relatedness fall within the narrow range of 0.0–0.25, as opposed to 0.50 for parent-offspring pairs and full-sibs, or 1.0 for clonemates). Now, however, with potential access to thousands or tens of thousands of SNPs and other genetic variants per specimen, it will be worthwhile to explore in detail the degree to which large numbers of unlinked genetic markers might be employed to estimate genome-wide kinship between pairs of individuals in local demes. Such estimates first should be

“ground-truthed”, and a good starting point will be to compare marker-based estimates of genetic relatedness with known levels of kinship in demes with well-established pedigrees (an approximate example of this approach, described at the Trondheim symposium, is provided by Bömcke and Gengler 2009). If it does generally prove feasible to obtain precise and reliable kinship estimates from multitudinous marker loci, tremendous new research opportunities would arise. For example, it would become possible, for the first time, to analyze possible correlations in nature between various social behaviors and kinship for individuals with specifiable coefficients of relative genetic relatedness. The fields of behavioral genetics and sociobiology (as well as conservation genetics) could be greatly enriched by this genomics-based capacity to quantify kinship precisely.

A second arena in which the genomics revolution should greatly expand inventorial capacity is in the field of phylogeography (Avice 2000), which to date has relied disproportionately on gene trees provided by mitochondrial (mtDNA) haplotypes. Because any single gene tree (nuclear or mitochondrial) provides only a tiny and potentially misrepresentative sample of an organism’s composite genealogical history (Degnan and Rosenberg 2009), geneticists have long sought to characterize nuclear gene trees to complement those from mtDNA (Palumbi and Baker 1994). Although some progress has been made (Hare 2001; Machado and Hey 2003), daunting hurdles remain, including the general technical challenge of isolating nuclear haplotypes from diploid specimens, and finding nuclear loci with relatively low levels of intra-genic recombination yet high levels of nucleotide sequence diversity. [The latter problem introduces a potential Catch-22, because sequence diversity and recombination rate appear to be positively correlated in at least some species (Nachman 2001; Lercher and Hurst 2002).] The genomics revolution will open new avenues for exploring how to generate nuclear gene trees. For example, it will permit richer characterizations of variation in recombination rates across the nuclear genome, greatly expand the numbers of candidate nuclear loci from which haplotype trees might be extracted, and in general improve our understanding of nuclear genome architecture in ways that should inform attempts to gather genealogical data from particular genomic regions.

With regard to the isolation of nuclear haplotypes, one under-explored possibility would be to take advantage of nature’s own haplotype-producing mechanism: gametogenesis. If researchers can develop straightforward protocols for isolating and sequencing nuclear haplotypes from single gametes in any taxa, this would open a wealth of novel research opportunities. This gamete-based approach to population genetics should be technologically feasible

with suitable effort; in the laboratory of Norman Arnheim, it has been implemented successfully for more than two decades with respect to genotyping single sperm cells in humans and mice (Li et al. 1988). As shown repeatedly by Arnheim’s group (e.g., Arnheim et al. 2007), many genetic insights can emerge from sequencing gametic haplotypes (as opposed to standard diploid genotypes, where cis versus trans phases cannot readily be distinguished in individuals that are heterozygous at multiple sites).

A third inventorial arena where the genomics revolution should pay important dividends has been termed phylogenomics (Philippe et al. 2005): the use of mass quantities of sequence data (or other large categories of genomic information) to reconstruct more robust species phylogenies than could be expected from traditional molecular data at one or a few loci. Much effort in the field of conservation biology is directed toward biodiversity assessment for purposes of setting conservation priorities, and one suggestion has been that each extant taxon’s phylogenetic distinctiveness (the amount of independent evolutionary history carried within its genome) should be included in the calculus of the planning process (May 1990; Vane-Wright et al. 1991; Faith 1992). To the extent that phylogenetic assessments should contribute to conservation planning (arguments for and against can be found in Purvis et al. 2005), then phylogenomics might play an expanding role in conservation genetics.

Synopsis

I previously defined conservation genetics broadly as “the study of genetic patterns or processes in any context that informs conservation efforts” (Avice 2008). Conservation genomics could be defined similarly, as the study of genomic patterns or processes in any context that informs conservation efforts. The only real distinction lies in the magnitude of molecular information made available by the genomics revolution. Traditionally, genetic analyses were based on only minuscule samples of the genome, but as we increasingly enter the genomics era, scientists will have routine access to far more genome-wide data than ever before. Conservation genomics will continue the well-established tradition of conservation genetics by providing conservation-relevant information in two general arenas: taking genetic inventories of the biological world (the primary topic of this essay), and addressing functional questions about genomic operations. Within the inventory realm, some of the applications of the genomics revolution will involve quantitative improvements in estimates of population genetic or evolutionary parameters of relevance to conservation, but others applications will be genuine qualitative breakthroughs. In this brief essay, I have

speculated on several of the research arenas that have at least the potential to be transformed by the newest genomic technologies. It is always dangerous to predict future developments in science, but it seems safe to conclude that the genomics era will have considerable impacts on how biological inventories, often with relevance to conservation efforts, are conducted.

Acknowledgements I would like to thank the organizers of the Trondheim meeting for the invitation, the participants for exciting discussions, and the ESF CONGEN Networking Programme for financial support.

References

- Arnheim N, Calabrese P, Tiemann-Boege I (2007) Mammalian meiotic recombination hotspots. *Annu Rev Genet* 41:369–399
- Avise JC (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, MA
- Avise JC (2001) Cytonuclear genetic signatures of hybridization phenomena: rationale, utility, and empirical examples from fishes and other aquatic animals. *Revi Fish Biol Fish* 10:253–263
- Avise JC (2004) *Molecular markers, natural history, and evolution*, 2nd edn. Sinauer, Sunderland, MA
- Avise JC (2006) *Evolutionary pathways in nature: a phylogenetic approach*. Cambridge University Press, Cambridge
- Avise JC (2008) The history, purview, and future of conservation genetics. In: Carroll SP, Fox CW (eds) *Conservation biology: evolution in action*. Oxford University Press, Oxford, pp 5–15
- Avise JC, Hamrick JL (eds) (1996) *Conservation genetics: case histories from nature*. Chapman & Hall, New York
- Bömcke E, Gengler N (2009) Combining microsatellite and pedigree data to estimate relationships among Skyros ponies. *J Appl Genet* 50:133–143
- Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol* 24:332–340
- Ehrlich PR, Hanski I (2004) *On the wings of checkerspots: a model system for population biology*. Oxford University Press, Oxford, UK
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1–10
- Frankham R, Ballou JD, Briscoe DA (2002) *Introduction to conservation genetics*. Cambridge University Press, Cambridge
- Freeland JR (2005) *Molecular ecology*. Wiley, West Sussex
- Gibson G, Muse SV (2009) *A primer of genome science*, 3rd edn. Sinauer, Sunderland, MA
- Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends Ecol Evol* 16:700–706
- Hillis DM, Mortiz C, Mable BK (1996) *Molecular systematics*, 2nd edn. Sinauer, Sunderland, MA
- Hubby JL, Lewontin RC (1966) A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* 54:577–594
- Hudson ME (2007) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol Ecol Notes*. doi:10.1111/j.1471-8286.2007.02019.x
- Kendal WS (2003) An exponential dispersion model for the distribution of human single nucleotide polymorphisms. *Mol Biol Evol* 20:579–590
- Lercher MJ, Hurst LD (2002) Human SNP variability and mutation rate are higher in regions of high recombination. *Trends Genet* 18:337–340
- Li HH, Gyllensten UB, Cui XF, Saiki RK, Erlich HA, Arnheim N (1988) Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 335:414–417
- Machado CA, Hey J (2003) The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proc R Soc Lond B* 270:1193–1202
- May RM (1990) Taxonomy as destiny. *Nature* 347:129–130
- Mitton JB (1997) *Selection in natural populations*. Oxford University Press, New York
- Mitton JB, Pierce BA (1980) The distribution of individual heterozygosity in natural populations. *Genetics* 95:1043–1054
- Nachman MW (2001) Single nucleotide polymorphisms and recombination rate in humans. *Trends Genet* 17:481–485
- Ouborg NJ (2009) Integrating population genetics and conservation biology in the era of genomics. *Biol Lett*. doi:10.1098/rsbl.2009.0590
- Ouborg NJ, Angeloni F, Vergeer P (2010) An essay on the necessity and feasibility of conservation genomics. *Conserv Genet* (in press)
- Palumbi SR, Baker SC (1994) Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Mol Biol Evol* 11:426–435
- Papanicolaou A, Joron M, McMillan WO, Blaxter ML, Jiggins CD (2005) Genomic tools and cDNA derived markers for butterflies. *Mol Ecol* 14:2883–2897
- Pertoldi C, Wójcik JM, Tokarska M, Kawałko A, Kristensen TN, Loeschcke V, Gregersen VR, Coltman D, Wilson GA, Randi E, Henryon M, Bendixen C (2010) Genome variability in the European and American bison detected using the BovineSNP50 Illumina Genotyping BeadChip. *Conserv Genet* (in press)
- Philippe H, Delsuc F, Brinkmann H, Lartillot N (2005) Phylogenomics. *Annu Rev Ecol Syst* 36:541–562
- Purvis A, Gittleman JL, Brooks T (2005) *Phylogeny and conservation*. Cambridge University Press, Cambridge, UK
- Schonewald-Cox CM, Chambers SM, MacBryde B, Thomas WL (1983) *Genetics and conservation*. Benjamin/Cummings, Menlo Park, CA
- Smith TB, Wayne RK (eds) (1996) *Molecular genetic approaches in conservation*. Oxford University Press, New York
- Vane-Wright RI, Humphries CJ, Williams PH (1991) What to protect: systematics and the agony of choice. *Biol Conserv* 55:235–254
- Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol Ecol* 17:1636–1647