

Genomics and the challenging translation into conservation practice

Authors: Affiliations provided at the end of the document

- 5 Aaron BA Shafer^{1*}, Jochen BW Wolf^{1*}, Paulo C Alves², Linnea Bergström¹, Michael W Bruford³, Ioana Brännström¹,
Guy Colling⁴, Love Dalén⁵, Luc De Meester⁶, Robert Ekblom¹, Katie D Fawcett⁷, Simone Fior⁸, Mehrdad Hajibabaei⁹,
Jason A Hill¹⁰, A Rus Hoezel¹¹, Jacob Hoglund¹, Evelyn L Jensen¹², Johannes Krause¹³, Torsten N Kristensen¹⁴, Michael
Krützen¹⁵, John K McKay¹⁶, Anita J Norman¹⁷, Rob Ogden¹⁸, E Martin Österling¹⁹, N Joop Ouborg²⁰, John Piccolo¹⁹,
Danijela Popovic²¹, Craig R Primmer²², Floyd A Reed²³, Marie Roumet⁸, Jordi Salmona²⁴, Tamara Schenekar²⁵, Michael
10 K Schwartz²⁶, Gernot Segelbacher²⁷, Helen Senn¹⁸, Jens Thaulow²⁸, Mia Valtonen²⁹, Andrew Veale¹², Philippine
Vergeer³⁰, Nagarjun Vijay¹, Carles Vilà³¹, Matthias Weissensteiner¹, Lovisa Wennerström¹⁰, Christopher W Wheat¹⁰,
Piotr Zieliński³²

* means these authors coordinated writing of the manuscript.

- 15 Corresponding author: aaron.shafer@ebc.uu.se

Abstract

The global loss of biodiversity continues at an alarming rate. Genomic approaches have been suggested as a promising tool for conservation practice, and we discuss how scaling-up to genome-wide inference can benefit traditional conservation genetic approaches and provide qualitatively novel insights. Yet, the generation of genomic data and subsequent analyses and interpretations are still challenging and largely confined to academic research in ecology and evolution. This generates a gap between basic research and applicable solutions for conservation managers faced with multifaceted problems. Before the real-world conservation potential of genomic research can be realized, we suggest that current infrastructures need to be modified, methods must mature, analytical pipelines need to be developed, and successful case studies must be disseminated to practitioners.

25

Conservation biology and genomics

Like most of the life sciences, conservation biology is being confronted with the challenge of how to integrate the collection and analysis of large-scale genomic data into its toolbox. Conservation biologists pull from a wide array of disciplines in an effort to preserve biodiversity and ecosystem services [1]. Genetic data have helped in this regard by detecting, for example, population substructure, measuring genetic connectivity, and identifying potential risks associated with demographic change and inbreeding [2]. Traditionally, conservation genetics (see Glossary) has relied on a handful of molecular markers ranging from a few allozymes to dozens of microsatellites [3]. But for close to a decade [4], genomics – broadly defined high-throughput sampling of nucleic acids [5] – has been touted as an important advancement to the field, a panacea of sorts for the unresolved conservation problems typically addressed with genetic data [6,7]. This transition has led to much promise, but also hyperbole, where concrete empirical examples of genomic data having a conservation impact remain rare.

Under the premise that assisting conservation of the world's biota is its ultimate purpose, the emerging field of conservation genomics must openly and pragmatically discuss its potential contribution towards this goal. While there are prominent examples where genetic approaches have made inroads influencing conservation efforts (e.g., Florida panther augmentation [8,9]) and wildlife enforcement (i.e., detecting illegal harvest [10]), it is not immediately clear that the conservation community and society more broadly have embraced genomics as a useful tool for conservation. Maintaining genetic diversity has largely been an afterthought when it comes to national biodiversity policies [11,12], and attempts to identify areas that might prove to be essential for conserving biological diversity rarely mention genomics (e.g. [13,14]). An obvious reason for this disconnect is that many of the pressing conservation issues (e.g., [15,16]) simply do not need genomics, but instead need political will.

The traditional use of genetic data in conservation biology has been historically demarcated into two interrelated areas [3]: i) understanding how evolutionary processes such as genetic drift, selection, and migration shape genetic and phenotypic variation of natural populations and determine population structure; and ii) more specifically, describing the effects of low effective population size on genetic variation and population viability. Nested within these are more general conservation issues, such as resolving taxonomic uncertainties, preserving local adaptation, and offsetting inbreeding depression (Table 1). Whether genome scale data can improve inferences within these two areas and better inform conservation initiatives is still not clear. Further, there are a plethora of uncertainties that practitioners need to be aware of, and considerable obstacles that need to be overcome, before genomics can make the transition to applied

conservation science. Many of the qualitatively novel aspects of genomic analyses, which include monitoring of epigenetic markers [17], environmental DNA approaches to assay species communities [18], and transcriptome assays [19], are still at an exploratory stage and are far from seeing use in real-world conservation issues.

60

In this essay, we focus first on how traditional applications of genetics in conservation can benefit from scaling-up to genome-wide data. In particular, we highlight two key areas that have received attention in the literature: identifying adaptive loci and the increased resolution afforded to genomic tools. Although the attention is warranted, part of the goal of this essay is to highlight the uncertainties associated with their potential application, and suggest how conservation practitioners should best deal with the uncertainties arising from the novel possibilities genomic data offer. We argue both of these areas are still largely an academic pursuit, and compared to conservation genetics, conservation genomics is far from seeing regular application. We discuss why the uptake of genomics in conservation has been, and will likely continue to be, difficult for practitioners and propose solutions to overcome the gap between basic genomics research and conservation practice.

70

Scaling-up: what can genomics do for conservation genetics?

From a conservation geneticist's stand point, the genomic approach offers a dramatic increase in the number of variable genetic markers used (generally single nucleotide polymorphisms: SNPs). This advancement should in principle improve the accuracy of estimating diversity and population demographic parameters of conservation relevance [6,20].

75 Whole genome sequencing, sequencing targeted portions of the genome [21], and direct genotyping of a random subset of the genome (generically referred to as genotyping-by-sequencing or GBS [22]) permits the assaying of genomic diversity at many orders of magnitude above the traditional number of pre-designed genetic markers. These approaches, GBS in particular, bypass the laborious process of marker characterization, primer development, and genotyping required for microsatellites, and can be compared and combined across labs (which is difficult for microsatellite loci).

80 Genomics further opens up the possibility to screen individuals and populations for adaptive loci, which is suggested by some to be the biggest potential contribution of genomics to conservation [5]. While targeting candidate genes (presumed adaptive) or serendipitously finding neutral markers linked to adaptive loci have not been uncommon in the conservation genetics era, given the low level of linkage disequilibrium often seen in natural populations [23] and polygenic nature of many traits [24], screening the entire genome holds considerably more power. Importantly, experimental systems have given us clues as to the signatures adaptive evolution leaves on the genome [25] and the academic interest in using genomics to explore local adaptation in the wild has grown considerably (e.g., [26,27]).

Conservation genomics is directly impacted as it draws from the methods and models developed in this basic research.

90 Numerous reviews have covered these advancements in relation to conservation in detail [6,20,28-33], and novel techniques and analyses are still emerging (Box 1). It is fair to say, however, that to date conservation genomics and the broader fields of ecological and evolutionary genomics are far from mature. While much praise and promise has come with adding genomics to the conservationist's toolbox, challenges and uncertainties that do not exist with conservation genetic data have also arisen.

95 *The detection of adaptive loci*

Nearly every review of conservation genomics has discussed the potential of emerging genomic tools to identify and study genetic variation with fitness consequences, which are potentially involved in local adaptation [6,20,28,33]. The inclusion of markers that reflect local adaptation would augment the identification of conservation units and improve detection of genomic regions causing inbreeding depression [6,28,34]. Identifying adaptive markers would also be
100 useful for conserving evolutionary processes (e.g., [35,36]), and genes associated with increased relative fitness could be propagated via human assistance, possibly including purposeful hybridization to maximise adaptability to changing environments [37,38]. This facet of conservation genomics clearly has the potential to provide information on the species, population, and individual level that was inaccessible using traditional genetic markers.

105 However, the detection of adaptive loci is challenging and interpretations are often vague. Screening genomes for loci under selection, let alone providing clear evidence for adaptive variation, is still an active area of academic debate and research [39,40]. It is becoming increasingly clear that many genes with small effects underlie phenotypic variation of fitness relevance [41,42], which makes their detection difficult. A further challenge is that the loci underlying adaptation need not exhibit clear signals of recent selection [43], and while studies have shown genomes under
110 pervasive selection with many loci linked to adaptive genetic variation [27], others have documented virtually homogenous genomes with only a few regions of presumed fitness significance [26]. Thus while genome-wide data provides access to the adaptive genomic landscape, the signatures of adaptation vary on case-by-case basis and could simply go undetected. In current practice, adaptive loci are often equated with outlier loci showing extreme levels of population differentiation [39,44]. Yet, outlier loci often fall within areas of low recombination and are often the result
115 of non-adaptive processes [39]; in such cases, the outlier regions could simply be reflective of demography, background selection in areas containing genomic features (i.e., centromeres), or sequence assembly artefacts. For example, in a

genomic study on flycatchers (*Ficedula* spp.) nearly all outlier peaks were likely not the result of selection imposed by the environment, but instead reflected areas of low diversity generated by linked purifying selection [39,45]. Signals, perhaps prematurely associated with local adaptation, could instead be caused by a number of other processes, such as mate choice [26], selfish alleles or background selection acting on a heterogeneous recombination landscape [46-49]. Many genome data sets of wild populations will have neither a genome assembly to assess spatial distribution of putative adaptive variation [50], nor the necessary information on recombination rates to avoid false positive and false negative inferences of local adaptation.

125 Given the multitude of uncertainties surrounding the detection and interpretation of adaptive loci, how can they best be implemented into a conservation framework? In his seminal essay, Michael Soulé [1] discussed the need for conservation biologists to tolerate some degree of uncertainty, and this will be the case for adaptive loci for the foreseeable future. While explicit conservation frameworks incorporating adaptive loci have been proposed [28], empirical examples and case studies of their application are still lacking. Where clear ecotypes exist, diagnostic outlier loci can be a valuable tool for monitoring stocks (e.g., salmon fisheries; [51]) regardless of their adaptive significance. In less clear cut cases the framework of Funk et al. [28], that advocates for a hierarchical approach using a combination of outlier and neutral loci to manage populations is a valid approach. As there is likely to be a trade-off between genetic uniqueness and genomic diversity [52,53], we do not advocate for managing only outlier loci; instead, where putative genetic contributions to adaptive variation can be identified, it should still be standard practice to incorporate additional information in the form of neutral markers or phenotypic data.

This still leaves a broader question unanswered: how much weight should be given to putatively adaptive variation that appears important at present, but holds no information about future adaptive potential? As the proportion of gene sequences with functional annotations grows, the value of individual loci will increase for conservation (e.g., detecting disease susceptibility [54]), and it is conceivable that loci of known function could be managed and propagated under certain conservation scenarios (e.g., disease outbreak in a small population - see also Box 2). Functional annotation can be borrowed from related species [55] and databases (e.g., Gene Ontology Consortium) under the assumption of orthology, though with unclear implications for the organism in question [56]. But caution is warranted here for one reason: using human genetics as the example, even with whole-genome sequencing efforts on large sample sizes, the success in uncovering the genetic basis of rare recessive diseases – and it is those that will be exposed by inbreeding – is still very limited [57]. Almost a decade of genomic work on the endangered California condor (*Gymnogyps*

californianus) [58] has yet to uncover the genetic basis for chondrodystrophy – a recessive and lethal form of dwarfism. These examples, combined with the above arguments, suggest that we should not expect, or invest a large effort, in detecting single adaptive genes of real consequence for conservation. Even if detected, the overall impact of individual loci on conservation and management will still remain tangential (e.g., through population models and projections [54]) until genotype-phenotype correlations of the focal species can be inferred with a high degree of certainty.

Increased population genomic resolution

In the realm of population genetics there are clear advantages that come with ample genome-wide data. Notably, genomic data should improve estimates of population demography [44] that are becoming increasingly amenable to model-based inference using Approximate Bayesian Computation or full likelihood models [59,60]. The characterization of processes such as inbreeding will also be more precise [34,61,62] and genomic data can be a useful monitoring tool [63] or reveal subtle population structure that is relevant for stock management (Box 3). Novel analytical approaches using haplotype information [64] or long stretches of homozygous sites [65] to infer population demographic histories have produced particularly exciting results. But similar to adaptive loci, using genomic data to infer demographic parameters is still very much an area of active academic research [66,67]. User friendly software that accommodates genomic data – a key for practical application – is still limited (e.g. [67-69]), and high performance computer clusters are generally required for data storage and analysis. Demographic studies that have explicitly tested the utility of genomic data have shown that essentially whole genome sequencing should be used [60], estimates can be biased [39,70,71], and some parameters simply cannot be estimated [71]. While genomics does offer increased resolution to addressing many of the outstanding conservation genetic problems (Table 1), we must still accept some, perhaps large, degree of uncertainty in the estimates produced.

Here, we propose that conservation practitioners are best served by focussing on broad scale population genetic patterns that might hold relevance to conservation issues of interest. From a practical standpoint the difference between 3 and 5 migrants per generation is not important, but 3 versus 500 is. Explicit formulation of the role genomics can play in achieving conservation priors – meaning a predetermined objective aimed at improving population viability – is a useful model in this regard [5]. If maintaining genetic connectivity were set as a conservation prior for example, differentiating low versus high migration scenarios with genomic data would clearly be informative. We stress, however, that even with more precise estimates the organism's biology still must be taken into account. If there is undetected family or population substructure represented in the genomic data, the estimated demographic parameters, which generally rely

on coalescent models, will reflect some hierarchical level (i.e. the region or species) instead of the population of interest at its current state. This effect is illustrated in a conservation study on Chinook salmon (*Oncorhynchus tshawytsca*), where effective population size estimates inferred with genomic data were occasionally higher than the census size [70].
180 This is a conceptual issue that also influences conservation genetics, but is an important consideration that will impact the utility of the parameters estimated from genome-scale data that might otherwise be overlooked by naively expecting higher accuracy.

The gap between genomics research and conservation application

185 We are approaching the \$1,000 genome [72], meaning that generating genomic information has become increasingly accessible even for non-model organisms with large genome sizes [50]. The pursuit of improved resolution and adaptive loci will transform many conservation genetics labs. Additionally, as life science companies dismantle old instrumentation, discontinue software, and support only the new high-throughput instruments, many conservation genetic applications will simply be forced into using genome-wide approaches [5]. At the same time, significant gaps
190 remain in the transfer of fundamental genomic research to end-user conservation application. The scientific and policy-practitioner communities operate in largely separate spheres (Figure 2), and we (the authors) feel that introducing genomics into the equation will increase the gap. In the pre-genomics era of microsatellites and Sanger sequencing, individual labs could start with sample collection and proceed all the way to data analysis and application. With next-generation sequencing, data generation is predominately outsourced and data cleaning and analysis often requires expert
195 consultation, creating multiple layers that simply do not exist with conservation genetics. Thus, advances in genomic methods may in fact contribute to an increasing gap between research and application without a concerted effort on the part of both scientists and conservation practitioners to build effective bridges. Broadly speaking, these gaps can be described in terms of the knowledge, tools (i.e., user friendly pipelines), finances, and communications needed to link fundamental research with applied science. Funding to the basic and fundamental research institutions is often not
200 compatible with frontline conservation actions, and will only aid in biodiversity conservation if applied research – which is often seen as repetitive and incremental from a basic research perspective – is supported (Figure 2).

Bridging the conservation genomics gap

In our view, it seems most fruitful in the current academic and conservation frameworks to encourage genome scientists
205 to communicate their research to practitioners and stakeholders and apply their techniques to conservation issues where appropriate. We stress that, for the conservation practitioner, it is not important to differentiate genetic from genomic

methods; rather it is sufficient to convey that we can now address a broader range of questions than before. The uncertainty in the new approaches should be clearly acknowledged. In the long run, it will be crucial to develop a range of laboratory protocols (e.g., [73]) and fail-safe tools that support conservation genomic applications and deliver tangible benefits to conservation managers (Box 3).

However, there is a more systemic problem with the current state of conservation genomics, in that there is little incentive for academic researchers – whom in many ways lead the conceptual debate and the development of genomic tools vital to application – to engage fully in applied conservation. Applied conservation genomics research is generally not reinforced in current funding schemes, and some academic research is branded as conservation (perhaps only as a selling feature for publication), even when it holds little real-world conservation value. Developing a genomic tool or framework that can be implemented by practitioners requires a level of rigor and repetition that is not conducive in the 'publish or perish' climate. Thus, we must rethink how the academic and conservation community funds conservation genomics research. In particular, we envision a research-policy framework analogous to translational medicine, or colloquially “bench-to-bedside”, that enables basic research to have an applied impact [74]. Here, the funding is partly driven by designated conservation questions, and application and uptake are the measured currency, not (just) publications (Figure 2). The applied output, evaluated by experts, leads directly into policy and ultimately cycles back to the funding body. Emerging examples, like the UK's research evaluation framework that specifically evaluates impact beyond academia, and Sweden's split governmental funding bodies with one emphasizing a link to stakeholders, should encourage bridging of the gap.

The key for conservation genomics will be for the academic and policy spheres to communicate in an effort to maintain a firm grasp on conceptual advancements (driven by academic research) and on-site conservation needs (recognized by practitioners). We are not so naive as to think that we can restructure the current academic funding and political climate, but recognizing how the current system fails to deliver tangible results is a starting point. Identifying the critical gains from genomic analysis in relation to drivers like conventions (e.g., Convention on International Trade in Endangered Species: CITES) or extinction risks (e.g., International Union for Conservation of Nature (IUCN) Red List) will be vital in this regard. In other words, if the IUCN or CITES request specific information to inform their conservation decisions, can genomics be the tool to provide it? We would be remiss not to note that some progress has been made bridging these barriers. The IUCN recently formed a Conservation Genetics Specialist Group that evaluates such critical gains, and over the last three years the EU Framework 7 Support Action Project ‘ConGRESS’ [75,76] has engaged over 300

stakeholders in Europe and canvassed their opinions on the use of genetics in practical conservation. While a need for the application of genetic technologies could be demonstrated to ConGRESS, a clear link to genomics was not yet evident. Thus, conservation genomics still must prove its worth; explicit examples need to be disseminated to the conservation community of realized gains from genomics.

Genomics will not solve many of the current conservation problems simply because the resolution of these problems depends on societal dynamics and political will. Genomics is also not a substitute for biological replication or good experimental design; more sequencing will not provide an answer *per se*, and the resolution required to address the question at hand (or conservation prior) needs serious consideration. Still, we believe that genomics can play an integral part in resolving conservation issues. With time will come matured methods, supported pipelines, and hopefully an integrated research-to-application framework that will bridge the conservation genomics gap. We encourage both applied and basic genome scientists to apply their work to conservation issues with the hope of alleviating some of the uncertainties discussed in this essay, and ultimately provide clear examples of conservation genomics in action.

250

Acknowledgements

This manuscript arose from the workshop 'Conservation Genomics: Academic exercise or transition with real-world implications?' organized by ABAS and JBWW held in Wik's Castle, Uppsala, Sweden from 18-21 March 2014. The workshop was funded by the ConGenOmics Initiative of the European Science Foundation (Refnr. 5005), the Swedish Research Council (ID: 70720201), and Uppsala University. The workshop program has been included as supplementary material (Appendix S1). All authors vividly contributed to the discussions and participated in drafting the manuscript. The final version of the manuscript benefited from comments by Paul Craze, Paul Hohenlohe, and an anonymous reviewer.

Author affiliations

1. Uppsala University, Ecology and Genetics, Evolutionary Biology Centre, 75236, Sweden
2. Universidade do Porto, CIBIO/InBIO & Faculdade de Ciências, 4485-661, Portugal
3. Cardiff University, School of Biosciences, CF10 3AX, United Kingdom
4. Musée national d'histoire naturelle Luxembourg, Population Biology, 2160, Luxembourg
5. Swedish Museum of Natural History, Bioinformatics and Genetics, 10405, Sweden
6. KU Leuven - University of Leuven, Aquatic Ecology, Evolution, and Conservation, 3000, Belgium
7. University of Groningen, Behavioural Ecology and Self-organization, 9712, Netherlands
8. ETH Zurich, Integrative Biology, 8092, Switzerland
9. University of Guelph, Integrative Biology, N1G 2W1, Canada
10. Stockholm University, Zoology 106 91, Sweden
11. Durham University, Biological and Biomedical Sciences, DH1 3LE United Kingdom
12. University of British Columbia Okanagan, Biology, V1V 1V7, Canada
13. University of Tuebingen, Archaeological Sciences, 72070, Germany
14. Aalborg University, Biotechnology, Chemistry and Environmental Engineering, 9220, Denmark
15. University of Zurich, Anthropological Institute & Museum, 8057, Switzerland
16. Colorado State University, Bioagricultural Sciences and Pest Management, 80523-1177, United States
17. Swedish University of Agricultural Sciences, Wildlife, Fish and Environmental Studies, 901 83 Sweden
18. Royal Zoological Society of Scotland, WildGenes Laboratory, EH12 6TS, United Kingdom
19. Karlstad University, Biology, 651 88, Sweden
20. Radboud University Nijmegen, Experimental Plant Ecology, 6500GL, Netherlands

21. University of Warsaw, Centre of New Technologies, 00-681, Poland
22. University of Turku, Biology, 20014, Finland
- 285 23. University of Hawai'i at Manoa, Biology, 96822, United States
24. Instituto Gulbenkian de Ciência, Population and conservation genetics group, 2780-156, Portugal
25. Karl-Franzens University Graz, Zoology, 8010, Austria
26. USDA Forest Service, Rocky Mountain Research Station, 59801, United States
27. University Freiburg, Wildlife Ecology and Management, 79106, Germany
- 290 28. Norwegian Institute for Water Research, Freshwater Biology, N-0349, Norway
29. University of Eastern Finland, Biology, 80101, Finland
30. Wageningen University, Nature Conservation and Plant Ecology, 6708 PB, Netherlands
31. Estación Biológica de Doñana, Conservation and Evolutionary Genetics Group, 41092, Spain
32. Jagiellonian University, Institute of Environmental Sciences, 30-387, Poland

295

Table 1. Main areas traditionally addressed by conservation genetics [3], current status of genetic and genomic approaches and the contribution genomics can potentially make. The far-right column details steps required before genomic research can regularly be applied to conservation issues.

Category	Status of conservation genetics	Possible contribution of conservation genomics	Required for transition from basic to applied
i) Evolutionary genetics of natural populations			
Demographic inference – population history	<input type="checkbox"/> Regularly used <input type="checkbox"/> Moderate resolution	<input type="checkbox"/> Improved accuracy and precision <input type="checkbox"/> Finer-scale population structure <input type="checkbox"/> Less limited by sample size	<input type="checkbox"/> Clear understanding of biases <input type="checkbox"/> User-friendly software
Adaptive genetic variation	<input type="checkbox"/> Minimally used <input type="checkbox"/> Limited to population correlations [77] or candidate gene approaches	<input type="checkbox"/> Improved detection of adaptive loci <input type="checkbox"/> Management frameworks proposed [28] <input type="checkbox"/> Methods still emerging and interpretations unclear	<input type="checkbox"/> In-depth validation studies <input type="checkbox"/> Genome annotation
Quantitative genetic variation	<input type="checkbox"/> Limited resolution <input type="checkbox"/> Often dependent on pedigrees or targeted gene approaches	<input type="checkbox"/> Improved detection of quantitative trait loci <input type="checkbox"/> Active application (e.g., genome-wide association studies)	<input type="checkbox"/> Ecological studies <input type="checkbox"/> Genome annotation <input type="checkbox"/> User-friendly pipelines
Taxonomic identification and general diagnostics	<input type="checkbox"/> Regularly used <input type="checkbox"/> Moderate resolution <input type="checkbox"/> Restricted to single individuals	<input type="checkbox"/> Assay species simultaneously [78] <input type="checkbox"/> Improved hybridization detection <input type="checkbox"/> Improved detection of pathogens	<input type="checkbox"/> Defined pipelines (Box 3) <input type="checkbox"/> Repeatability
ii) Effects of small population size			
Inbreeding detection	<input type="checkbox"/> Regularly used <input type="checkbox"/> Limited resolution [34]	<input type="checkbox"/> Improved estimates of inbreeding [34,62] <input type="checkbox"/> Novel genomic metrics [79] <input type="checkbox"/> Assess impact on specific genomic regions or adaptive loci	<input type="checkbox"/> User-friendly bioinformatics <input type="checkbox"/> Genome annotation <input type="checkbox"/> Practitioner demand
Population viability	<input type="checkbox"/> Minimally used [80]	<input type="checkbox"/> Improved estimates of inbreeding metrics used in viability models [80]	<input type="checkbox"/> Practitioner demand
Additional applications			
Genetic monitoring	<input type="checkbox"/> Minimally used [11]	<input type="checkbox"/> Improved sampling regimes [63] <input type="checkbox"/> More powerful biodiversity surveys	<input type="checkbox"/> Practitioner demand <input type="checkbox"/> Government compliance [11]
Population census	<input type="checkbox"/> Regularly used	<input type="checkbox"/> Limited need for application	<input type="checkbox"/> Practitioner demand
Maternity, paternity, and kinship analysis	<input type="checkbox"/> Regularly used	<input type="checkbox"/> Useful when microsatellite power is limited [81]	<input type="checkbox"/> Practitioner demand

Glossary Box

- 305 **Adaptive locus:** a region of the genome under selection that codes for (or is closely linked to a causative locus) a phenotype with fitness consequences in a particular environment.
- Annotation:** the process of delineating and assigning function to genetic sequences.
- 310 **Background selection:** the loss of genetic diversity at neutrally evolving sites that are linked to sites under purifying selection.
- Candidate genes:** genes putatively underlying variation in a certain phenotype.
- 315 **Coalescent theory:** a retrospective population genetic framework that traces genetic variants of a locus to the most recent common ancestor. Used to infer demographic parameters of population histories.
- Conservation genetics:** uses genetic markers to conserve biodiversity and help manage species or populations. Traditional genetic markers include: allozymes, microsatellites, and targeted gene sequence.
- 320 **Conservation genomics:** uses genome-wide information to conserve biodiversity and help manage species or populations. Genomic data is derived from high-throughput sequencing technology. Relevant examples are whole genome re-sequencing and targeted approaches like exome sequencing, genotyping-by-sequencing, SNP genotyping, and transcriptome sequencing.
- 325 **Effective population size:** a population genetic convention describing the number of breeding individuals in an ideal population that would lose genetic variation at the same rate as the observed population.
- Environmental DNA:** is DNA found in environmental samples (e.g., water, soil) that can be used in genetic or genomic analysis. This contrasts traditional approaches that target a specific organism or tissue.
- 330 **Genetic drift:** the loss of genetic variants due to random sampling from one generation to the next.
- Genome assembly:** the process of ordering and orienting sequencing into a contiguous consensus sequence of the genome.
- 335 **Genotyping-by-sequencing:** the sequencing of a repeatable subset of the genome seeded by restriction enzyme recognition sites. Restriction-site Associated DNA sequencing or RAD-seq is another commonly used term.
- 340 **Haplotypes:** a particular combination of alleles at collinear positions along a stretch of DNA.
- Inbreeding:** the increase of genomic segments in identity-by-descent due to mating between closely related individuals. Results in an increase in homozygosity potentially revealing detrimental recessive alleles with negative fitness consequences.
- 345 **Linkage disequilibrium:** the non-random association of alleles at two or more loci.
- Outlier locus:** a region of the genome that, based on user-defined criteria (often extreme population differentiation), deviates from the rest of the entire genome.
- 350 **Orthology:** homologous DNA sequence descended from a shared common ancestor
- Recombination:** the process of genetic exchange between homologous chromosomes, often resulting in a new combination of alleles.
- 355 **Transcriptome:** set of all RNA molecules transcribed from a DNA template.

360 **Figure 1 (in Box 2).** Genetic rescue on a population of (a) Bighorn sheep (*Ovis canadensis*) at National Bison Range, Montana, USA. Two separate introductions successfully increased genetic diversity and reversed the negative effects of inbreeding [82]. Miller et al. [83] used population monitoring and genome-scale data to document the demographic impact of the introductions. (b) The hybrid index based on genome-wide data showed that migrant alleles (from translocated individuals) increased over time, and life history traits including (c) longevity and (d) number of offspring
365 both increased with migrant ancestry.

Figure 2. Schematic for the translation of conservation genomics research to conservation policy and action. Currently there are two largely separate spheres of applied and basic research. The black lines are the basic research to applied
370 workflow with the green lines reflecting the extent to which conservation genetics and genomics currently fit into this scheme – note that conservation genetics has integrated into the applied sphere. The dashed green line represents the gap we see between the academic groundwork readily embracing genomic technology and on-site conservation needs. The grey lines are reflective of the larger framework we envision (and that is emerging) where conservation questions directly fund conservation genomic research and feed into management and biodiversity policies. Importantly, the
375 applied component maintains a firm link to the conceptual advancements driven by basic research.

Box 1 – An emerging area stemming from ancient DNA technology

380 Genomic analysis of ancient samples – considered anything from a museum specimen to archaeological specimens [84] – can establish baseline levels of genetic parameters in ancestral populations prior to demographic declines [85]. This concerns both the amount of genetic variation that might have been lost over time and the origin of contemporary population structure. The timing of population fragmentation and how these are related to past changes in the environment (e.g., anthropogenic impact or climate change) can provide valuable insight into current processes influencing population viability. For example, whether low effective size or high population structure is inherent to the biology of a species or a direct result of recent anthropogenic disturbance is useful information to have when factoring genomic data into conservation planning.

385

Ancient DNA and genomic approaches provide another advantage in that they are less constrained by sample quality. The latter is illustrated by the fact that entire genomes have been reconstructed, using optimized genomic approaches, from 400,000 year old samples of early hominids [86]. Sample quality is of direct concern to conservation biology, since sample collection and quality become increasingly difficult when the species of interest is rare or elusive in the wild. Conservation biologists often need to resort to museum samples or non-invasive sampling (e.g. scat, hair, scales) of poor quality. Utilizing genomic techniques often employed for ancient DNA studies (e.g., [85,87]), holds potential not only to recover genetic information from the species of interest, but additionally reveal aspects of pathogens (e.g., [88]) that might prove relevant for conservation and management.

395 **Box 2 Conservation genomics and translocations**

Endangered species are often vulnerable to pathogen outbreaks both in captive breeding populations and in small, isolated natural populations. With high throughput sequencing approaches even trace amounts of pathogen or parasite DNA can be detected from various types of samples (e.g., environmental DNA [18]) and host species can be readily identified [89]. This technology provides a rapid and cost efficient way to identify and monitor pathogen load in 400 populations at risk. Translocation and re-introduction programs could benefit from rapid genomic testing of pathogen load prior to release and select individuals based on desired criteria. Genomic screening could also allow for the selection of individuals according to specific genetic signatures or simply maximum variation. For the purpose of conservation interventions such as the founding of captive populations, performing genetic rescues (i.e., restoring positive growth [90]), or assisting migration, these data could be vital and the ability to monitor the outcomes after such 405 interventions would be enhanced (e.g., Figure 1; [83]).

Box 3 – From the lab to law: FishPopTrace

The step change afforded by genomic techniques and the transition from fundamental research to application can both be demonstrated by the development of genomic tools for fisheries management in Europe. The FishPopTrace project funded by the European Union Framework Program 7 utilized genome-wide technologies to identify and genotype SNP markers in four species of commercial fish: cod, herring, hake and sole. These data were used by fishery geneticists to describe populations within species and subsequently to identify minimum SNP panels of maximum power to trace the geographic origin of fish in trade. FishPopTrace revealed population structure at a geographic scale not previously recognized, leading directly to tools for applied conservation (e.g., identification of illegal trade and mislabelling). The UK government and the Marine Stewardship Council are now using the stock traceability information in a regulatory and authenticity verification framework. The challenge for applied conservation labs will be to develop similar pipelines and apply this level of rigour to less charismatic or economically valued species.

References

- 1 Soulé, M.E. (1985) What Is Conservation Biology? *BioScience* 35, 727–734
- 2 Frankham, R. (1995) Conservation genetics. *Annu. Rev. Genet.* 29, 305–327
- 420 3 Frankham, R. *et al.* (2001) *Introduction to Conservation Genetics*, Cambridge University Press.
- 4 DeSalle, R. and Amato, G. (2004) The expansion of conservation genetics. *Nat. Rev. Genetics* 5, 702–712
- 5 McMahon, B.J. *et al.* (2014) How and why should we implement genomics into conservation? *Evol. Appl.* DOI: 10.1111/eva.12193
- 6 Allendorf, F.W. *et al.* (2010) Genomics and the future of conservation genetics. *Nat. Rev. Genetics* 11, 697–709
- 425 7 Frankham, R. (2010) Where are we in conservation genetics and where do we need to go? *Conserv. Genet.* 11, 661–663
- 8 U.S. Fish and Wildlife Service (2008) *Florida panther recovery Plan (Puma concolor coryi)*, Third Revision, U.S. Fish and Wildlife Service.
- 9 Seal, U.S. and Lacy, R.C. (1994) *A plan for genetic restoration and management of the Florida panther (Felis concolor coryi)*, IUCN.
- 430 10 Manel, S. *et al.* (2002) Detecting wildlife poaching: identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. *Conserv. Biol.* 16, 650–659
- 11 Laikre, L. (2010) Genetic diversity is overlooked in international conservation policy implementation. *Conserv. Genet.* 11, 349–354
- 435 12 Laikre, L. *et al.* (2010) Neglect of genetic diversity in implementation of the convention on biological diversity. *Conserv. Biol.* 24, 86–88
- 13 Sutherland, W.J. *et al.* (2010) A horizon scan of global conservation issues for 2010. *Trends Ecol. Evol.* 25, 1–7
- 14 Carroll, S.P. *et al.* (2014) Applying evolutionary biology to address global challenges. *Science* DOI: 10.1126/science.1245993
- 440 15 Ripple, W.J. *et al.* (2014) Status and ecological effects of the world's largest carnivores. *Science* 343, 1241484–1241484
- 16 Fyumagwa, R. *et al.* (2013) Roads as a threat to the Serengeti ecosystem. *Conserv. Biol.* 27, 1122–1125
- 17 Danchin, É. *et al.* (2011) Beyond DNA: integrating inclusive inheritance into an extended theory of evolution. *Nat. Rev. Genetics* 12, 475–486
- 445 18 Taberlet, P. *et al.* (2012) Environmental DNA. *Mol. Ecol.* 21, 1789–1793
- 19 Vandersteen Tymchuk, W. *et al.* (2010) Conservation genomics of Atlantic salmon: variation in gene expression between and within regions of the Bay of Fundy. *Mol. Ecol.* 19, 1842–1859
- 20 Harrisson, K.A. *et al.* (2014) Using genomics to characterize evolutionary potential for conservation of wild populations. *Evol. Appl.* DOI: 10.1111/eva.12149
- 450 21 Mamanova, L. *et al.* (2010) Target-enrichment strategies for next-generation sequencing. *Nat. Methods* 7, 111–118
- 22 Narum, S.R. *et al.* (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Mol. Ecol.* 22, 2841–2847
- 23 Poelstra, J.W. *et al.* (2013) An extensive candidate gene approach to speciation: diversity, divergence and linkage disequilibrium in candidate pigmentation genes across the European crow hybrid zone. *Heredity* 111, 467–473
- 455 24 Robinson, M.R. *et al.* (2014) Explaining additional genetic variation in complex traits. *Trends Genet.* 30, 124–132
- 25 Burke, M.K. (2012) How does adaptation sweep through the genome? Insights from long-term selection experiments. *Proc. Roy. Soc. B* 279, 5029–5038
- 26 Poelstra, J.W. *et al.* (2014) The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* 344, 1410–1414
- 460 27 Soria-Carrasco, V. *et al.* (2014) Stick insect genomes reveal natural selection's role in parallel speciation. *Science* 344, 738–742
- 28 Funk, W.C. *et al.* (2012) Harnessing genomics for delineating conservation units. *Trends Ecol. Evol.* 27, 489–496
- 29 Steiner, C.C. *et al.* (2013) Conservation genomics of threatened animal species. *Annu. Rev. Anim. Biosci.* 1, 261–281
- 465 30 Kohn, M.H. *et al.* (2006) Genomics and conservation genetics. *Trends Ecol. Evol.* 21, 629–637
- 31 Angeloni, F. *et al.* (2011) Genomic toolboxes for conservation biologists. *Evol. Appl.* 5, 130–143
- 32 Primmer, C.R. (2009) From conservation genetics to conservation genomics. *Ann. N. Y. Acad. Sci.* 1162, 357–368
- 33 Ouborg, N.J. *et al.* (2010) Conservation genetics in transition to conservation genomics. *Trends Genet.* 26, 177–187
- 470 34 Miller, J.M. and Coltman, D.W. (2014) Assessment of identity disequilibrium and its relation to empirical heterozygosity fitness correlations: a meta-analysis. *Mol. Ecol.* 23, 1899–1909
- 35 Crandall, K. *et al.* (2000) Considering evolutionary processes in conservation biology. *Trends Ecol. Evol.* 15, 290–295
- 475 36 Hoffmann, A.A. and Willi, Y. (2008) Detecting genetic responses to environmental change. *Nat. Rev. Genetics* 9, 421–432

- 37 Barton, N.H. (2008) The role of hybridization in evolution. *Mol. Ecol.* 10, 551–568
- 38 Nolte, A.W. and Tautz, D. (2010) Understanding the onset of hybrid speciation. *Trends Genet.* 26, 54–58
- 39 Cruickshank, T.E. and Hahn, M.W. (2014) Reanalysis suggests that genomic islands of speciation are due to
480 reduced diversity, not reduced gene flow. *Mol. Ecol.* 23, 3133–3157
- 40 Vitti, J.J. *et al.* (2013) Detecting natural selection in genomic data. *Annu. Rev. Genet.* 47, 97–120
- 41 Santure, A.W. *et al.* (2013) Genomic dissection of variation in clutch size and egg mass in a wild great tit (*Parus major*) population. *Mol. Ecol.* 22, 3949–3962
- 42 Robinson, M.R. *et al.* (2013) Partitioning of genetic variation across the genome using multimarker methods in a
485 wild bird population. *Mol. Ecol.* 22, 3963–3980
- 43 Ågren, J. *et al.* (2013) Genetic mapping of adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 21077–21082
- 44 Luikart, G. *et al.* (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat. Rev. Genetics* 4, 981–994
- 490 45 Ellegren, H. *et al.* (2012) The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* 491, 756–760
- 46 Sandler, L. and Novitski, E. (1957) Meiotic drive as an evolutionary force. *Am. Nat.* XCI, 105–110
- 47 Charlesworth, B. *et al.* (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* 70, 155–174
- 495 48 Hu, X.-S. and He, F. (2005) Background selection and population differentiation. *J. Theor. Biol.* 235, 207–219
- 49 Stephan, W. (2010) Genetic hitchhiking versus background selection: the controversy and its implications. *Phil. Trans. R. Soc. B.* 365, 1245–1253
- 50 Ekblom, R. and Wolf, J.B.W. (2014) A field guide to whole genome sequencing, assembly and annotation. *Evol. Appl.* DOI: 10.1111/eva.12178
- 500 51 Russello, M.A. *et al.* (2011) Detection of outlier loci and their utility for fisheries management. *Evol. Appl.* 5, 39–52
- 52 Coleman, R.A. *et al.* (2013) Balancing genetic uniqueness and genetic variation in determining conservation and translocation strategies: a comprehensive case study of threatened dwarf galaxias, *Galaxiella pusilla* (Mack) (Pisces: Galaxiidae). *Mol. Ecol.* 22, 1820–1835
- 505 53 Bonin, A. *et al.* (2007) Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. *Conserv. Biol.* 21, 697–708
- 54 Robinson, S.J. *et al.* (2012) Emerging prion disease drives host selection in a wildlife population. *Ecol Appl* 22, 1050–1059
- 55 Primmer, C.R. *et al.* (2013) Annotated genes and nonannotated genomes: cross-species use of Gene Ontology in ecology and evolution research. *Mol. Ecol.* 22, 3216–3241
- 510 56 Pavlidis, P. *et al.* (2012) A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. *Mol. Biol. Evol.* 29, 3237–3248
- 57 Cirulli, E.T. and Goldstein, D.B. (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genetics* 11, 415–425
- 515 58 Romanov, M.N. *et al.* (2006) Construction of a California condor BAC library and first-generation chicken–condor comparative physical map as an endangered species conservation genomics resource. *Genomics* 88, 711–718
- 59 Gutenkunst, R.N. *et al.* (2009) Inferring the joint demographic history of multiple populations from multidimensional snp frequency data. *PLoS Genet* 5, e1000695
- 520 60 Robinson, J.D. *et al.* (2014) ABC inference of multi-population divergence with admixture from unphased population genomic data. *Mol. Ecol.* 23, 4458–4471
- 61 Hoffman, J.I. *et al.* (2014) High-throughput sequencing reveals inbreeding depression in a natural population. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3775–3780
- 62 Kardos, M. *et al.* (2013) Evaluating the role of inbreeding depression in heterozygosity–fitness correlations: how useful are tests for identity disequilibrium? *Mol. Ecol. Res.* 14, 519–530
- 525 63 Hoban, S. *et al.* (2014) Comparative evaluation of potential indicators and temporal sampling protocols for monitoring genetic erosion. *Evol. Appl.* DOI: 10.1111/eva.12197
- 64 Harris, K. and Nielsen, R. (2013) Inferring Demographic History from a Spectrum of Shared Haplotype Lengths. *PLoS Genet* 9, e1003521
- 530 65 Li, H. and Durbin, R. (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475, 493–496
- 66 Excoffier, L. *et al.* (2013) Robust demographic inference from genomic and SNP data. *PLoS Genet* 9, e1003905
- 67 Pinho, C. and Hey, J. (2010) Divergence with gene flow: models and data. *Annu. Rev. Ecol. Evol. Syst.* 41, 215–230
- 535 68 Excoffier, L. *et al.* (2005) Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* 169, 1727–1738

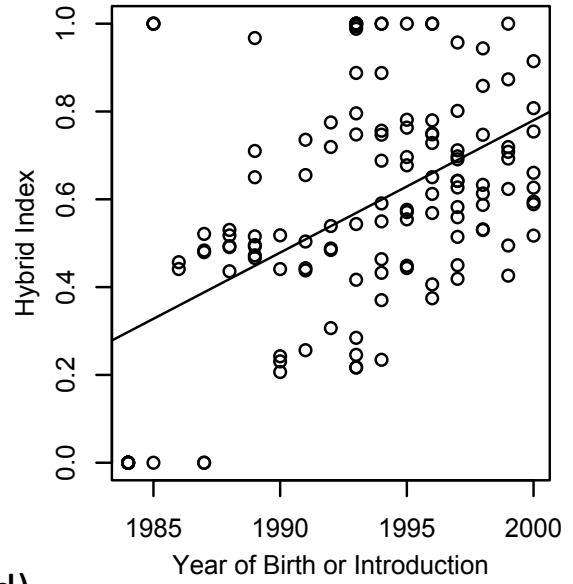
- 69 Cornuet, J.M. *et al.* (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences
about population history using single nucleotide polymorphism, DNA sequence and microsatellite data.
Bioinformatics 30, 1187–1189
- 540 70 Larson, W.A. *et al.* (2014) Genotyping by sequencing resolves shallow population structure to inform
conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evol. Appl.* DOI: 10.1111/eva.12128
- 71 Shafer, A.B.A. *et al.* Demographic inferences using short-read genomic data in an Approximate Bayesian
Computation framework: in silico evaluation of power, biases, and a proof of concept in Atlantic walrus. *Under
review* (included in supplementary material)
- 545 72 Hayden, E.C. (2014) Technology: The \$1,000 genome. *Nature* 507, 294–295
- 73 Zieliński, P. *et al.* (2014) Development, validation and high-throughput analysis of sequence markers in
nonmodel species. *Mol. Ecol. Res.* 14, 352–360
- 74 Woolf, S.H. (2008) The meaning of translational research and why it matters. *JAMA* 299, 211–213
- 75 Hoban, S.M. *et al.* (2013) Bringing genetic diversity to the forefront of conservation policy and management.
550 *Conserv. Genet. Resour.* 5, 593–598
- 76 Hoban, S. *et al.* (2013) Conservation genetic resources for effective species survival (congress): bridging the
divide between conservation research and practice. *J. Nat. Conserv.* 21, 433–437
- 77 Shafer, A.B.A. and Wolf, J.B.W. (2013) Widespread evidence for incipient ecological speciation: a meta-analysis
of isolation-by-ecology. *Ecology Letters*
- 555 78 Taberlet, P. *et al.* (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.*
21, 2045–2050
- 79 Polašek, O. *et al.* (2010) Comparative assessment of methods for estimating individual genome-wide
homozygosity-by-descent from human genomic data. *BMC Genomics* 11, 139
- 80 Frankham, R. *et al.* (2014) Genetics in conservation management: Revised recommendations for the 50/500 rules,
560 Red List criteria and population viability analyses. *Biol. Conserv.*
- 81 Tokarska, M. *et al.* (2009) Effectiveness of microsatellite and SNP markers for parentage and identity analysis in
species with low genetic diversity: the case of European bison. *Heredity* 103, 326–332
- 82 Hogg, J.T. *et al.* (2006) Genetic rescue of an insular population of large mammals. *Proc. Roy. Soc. B* 273, 1491–
1499
- 565 83 Miller, J.M. *et al.* (2012) Genomic consequences of genetic rescue in an insular population of bighorn sheep
(*Ovis canadensis*). *Mol. Ecol.* 21, 1583–1596
- 84 Pääbo, S. *et al.* (2004) Genetic analyses from ancient dna. *Annu. Rev. Genet.* 38, 645–679
- 85 Bi, K. *et al.* (2013) Unlocking the vault: next-generation museum population genomics. *Mol. Ecol.* 22, 6018–6032
- 86 Meyer, M. *et al.* (2013) A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* 505,
570 403–406
- 87 Perry, G.H. *et al.* (2010) Genomic-scale capture and sequencing of endogenous DNA from feces. *Mol. Ecol.* 19,
5332–5344
- 88 Bos, K.I. *et al.* (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human
tuberculosis. *Nature* DOI: 10.1038/nature13591
- 575 89 Thiemann, T.C. *et al.* (2011) Development of a high-throughput microsphere-based molecular assay to identify 15
common bloodmeal hosts of *Culex* mosquitoes. *Mol. Ecol. Res.* 12, 238–246
- 90 Carlson, S.M. *et al.* (2014) Evolutionary rescue in a changing world. *Trends Ecol. Evol.* 29, 521–530

Figure 1 in Box 2

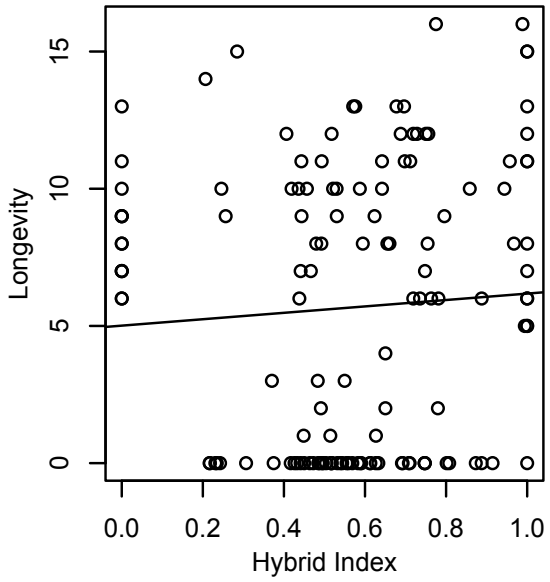
(a)



(b)



(c)



(d)

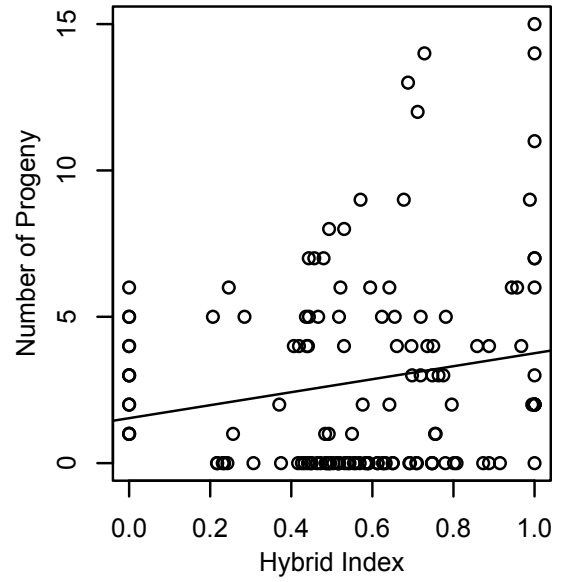
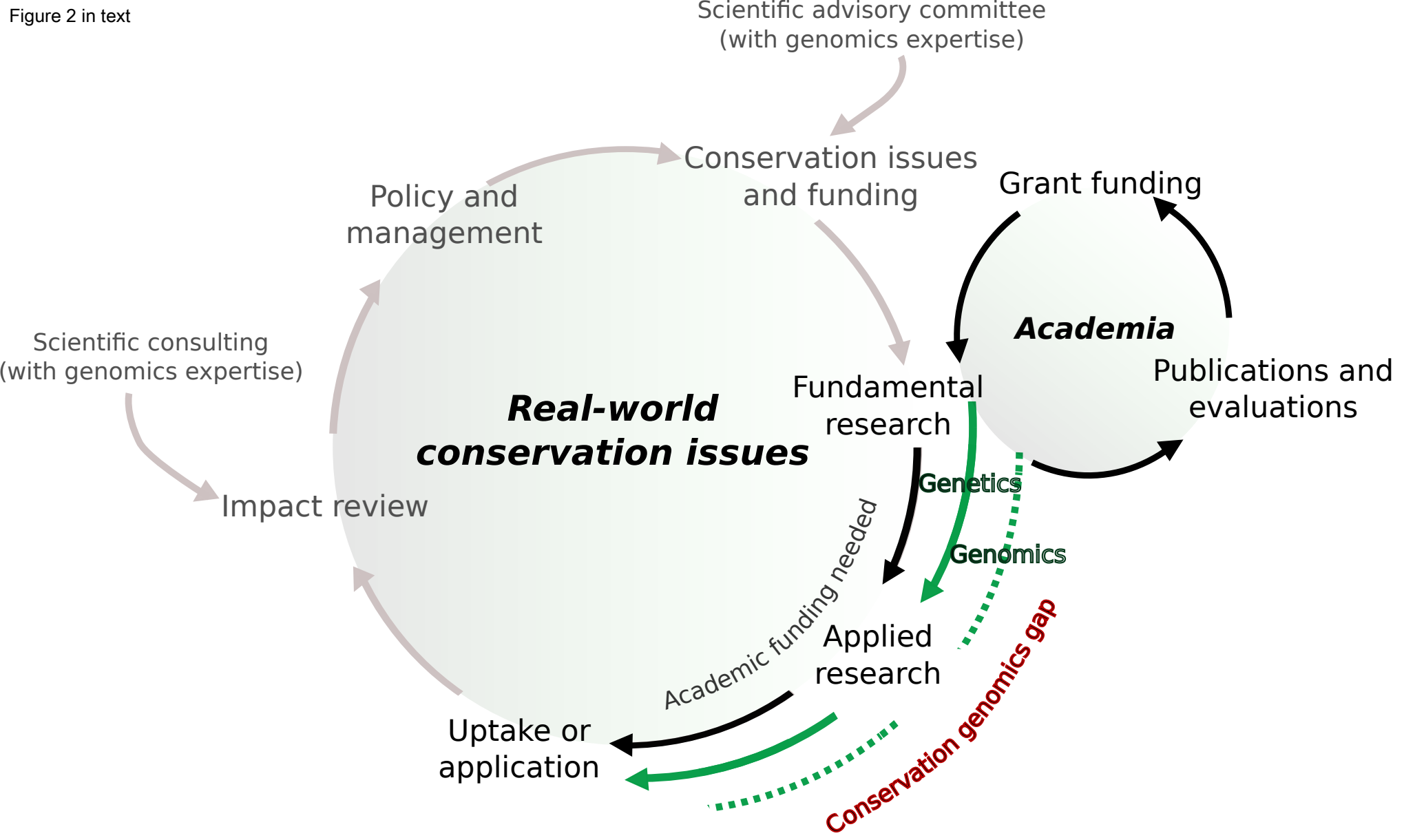


Figure 2 in text



Genomics research and development

SNP discovery

SNP validation and selection

Genome-wide genotyping

Marker assessment and selection

Population screening

Population genomic analysis

SNP panel selection

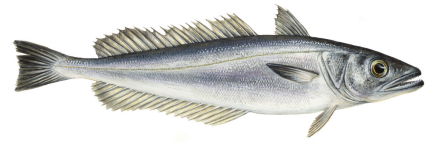
Applied traceability tools

Platform selection

Method validation

Standard Operating Procedures

The **FishPopTrace** target species
reproduced with permission from the Scandinavian Fishing Yearbook (c)



European hake (*Merluccius merluccius* L.)



Atlantic herring (*Clupea harengus* L.)



Atlantic cod (*Gadus morhua* L.)



Common sole (*Solea solea* L.)