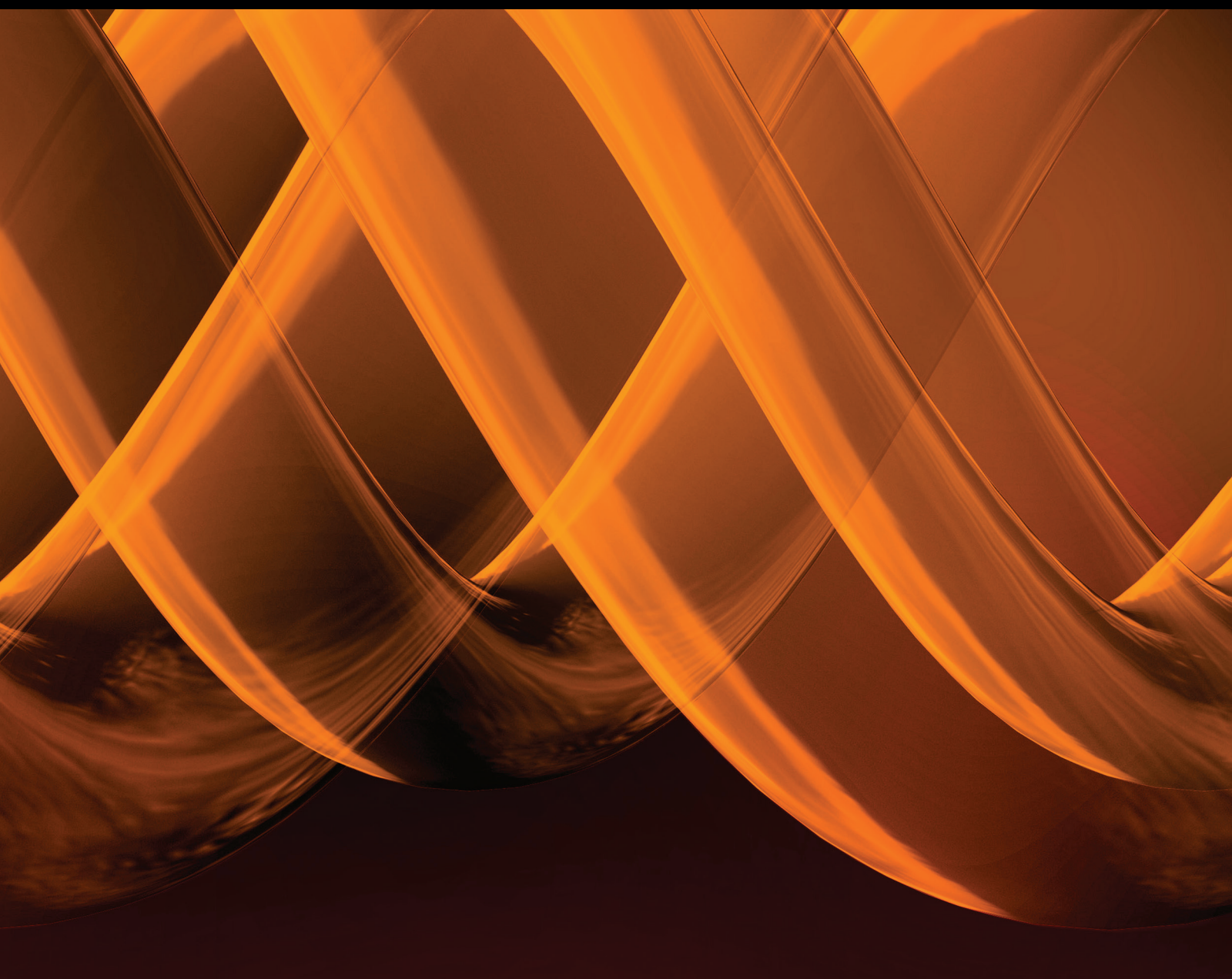
International Journal of Genomics

How Can Genomic Tools Contribute to the Conservation of Endangered Organisms

Guest Editors: Cino Pertoldi, Ettore Randi, Aritz Ruiz-González, Philippine Vergeer, and Joop Ouborg



How Can Genomic Tools Contribute to the Conservation of Endangered Organisms

International Journal of Genomics

How Can Genomic Tools Contribute to the Conservation of Endangered Organisms

Guest Editors: Cino Pertoldi, Ettore Randi, Aritz Ruiz-González, Philippine Vergeer, and Joop Ouborg

Editorial Board

Contents

How Can Genomic Tools Contribute to the Conservation of Endangered Organisms
Cino Pertoldi, Ettore Randi, Aritz Ruiz-González, Philippine Vergeer, and Joop Ouborg
Volume 2016, Article ID 4712487, 2 pages

Novel Graphical Analyses of Runs of Homozygosity among Species and Livestock Breeds
Laura Iacolina, Astrid V. Stronen, Cino Pertoldi, Małgorzata Tokarska, Louise S. Nørgaard, Joaquin Muñoz, Anders Kjærsgaard, Aritz Ruiz-Gonzalez, Stanisław Kami'nski, and Deirdre C. Purfield
Volume 2016, Article ID 2152847, 8 pages

Using Genome-Wide SNP Discovery and Genotyping to Reveal the Main Source of Population Differentiation in *Nothofagus dombeyi* (Mirb.) Oerst. in Chile
Rodrigo Hasbún, Jorge González, Carolina Iturra, Glenda Fuentes, Diego Alarcón, and Eduardo Ruiz
Volume 2016, Article ID 3654093, 10 pages

Integrating Genomic Data Sets for Knowledge Discovery: An Informed Approach to Management of Captive Endangered Species
Kristopher J. L. Irizarry, Doug Bryant, Jordan Kalish, Curtis Eng, Peggy L. Schmidt, Gini Barrett, and Margaret C. Barr
Volume 2016, Article ID 2374610, 12 pages

A Quantitative Genomic Approach for Analysis of Fitness and Stress Related Traits in a *Drosophila melanogaster* Model Population
Palle Duun Rohde, Kristian Krag, Volker Loeschcke, Johannes Overgaard, Peter Sørensen, and Torsten Nygaard Kristensen
Volume 2016, Article ID 2157494, 11 pages

The Microbiome of Animals: Implications for Conservation Biology
Simon Bahrndorff, Tibebu Alemu, Temesgen Alemneh, and Jeppe Lund Nielsen
Volume 2016, Article ID 5304028, 7 pages

Differential Methylation of Genomic Regions Associated with Heteroblasty Detected by M&M Algorithm in the Nonmodel Species *Eucalyptus globulus* Labill.
Rodrigo Hasbún, Carolina Iturra, Soraya Bravo, Boris Rebolledo-Jaramillo, and Luis Valledor
Volume 2016, Article ID 4395153, 7 pages

The Use of Genomics in Conservation Management of the Endangered Visayan Warty Pig (*Sus cebifrons*)
Rascha J. M. Nuijten, Mirte Bosse, Richard P. M. A. Crooijmans, Ole Madsen, Willem Schaftenaar, Oliver A. Ryder, Martien A. M. Groenen, and Hendrik-Jan Megens
Volume 2016, Article ID 5613862, 9 pages

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 4712487, 2 pages
http://dx.doi.org/10.1155/2016/4712487

Editorial

How Can Genomic Tools Contribute to the Conservation of Endangered Organisms

Cino Pertoldi,[1] Ettore Randi,[2] Aritz Ruiz-González,[3] Philippine Vergeer,[4] and Joop Ouborg[5]

[1]*Department of Chemistry and Bioscience, Aalborg University, Aalborg Zoo, Mølleparkvej 63, 9000 Aalborg, Denmark*
[2]*Istituto Superiore per la Protezione e la Ricerca Ambientale ISPRA, Roma, Italy*
[3]*Department of Zoology and Animal Cell Biology, University of the Basque Country, UPV/EHU, Vitoria-Gasteiz, Spain*
[4]*Nature Conservation and Plant Ecology Group, Wageningen University, Wageningen, Netherlands*
[5]*Experimentele Planten Ecologie, IWWR, Radboud Universiteit, Nijmegen, Netherlands*

Correspondence should be addressed to Cino Pertoldi; cp@bio.aau.dk

Received 19 October 2016; Accepted 19 October 2016

Conservation biologists have realized the urgent need for genomic tools and interdisciplinary approaches to help understand the loss of biodiversity. Conservation genetics has been successful in highlighting the roles of evolutionary and population genetics for biodiversity conservation, yet several critical issues remain. Rapid development of new sequencing techniques means conservation genomics can now help answer some of the crucial issues that conservation genetics was able to highlight but not to resolve. These include a better understanding of outbreeding and inbreeding depression and the extent to which estimates of heterozygosity accurately reflect quantitative trait variation and fitness. Conservation genomics can help resolve taxonomic uncertainties, contribute to designation of evolutionary significant units and ecotypes, and identify contemporary versus historic patterns of hybridization. The definition of a species is nevertheless a normative concept and cannot be resolved by genomics alone. All living organisms are faced with a multitude of challenges in their natural environment, such as climate change, diseases, predation, competition, and habitat disturbance. In the short term, animals and plants can acclimatize to shifting environmental conditions by developing and expressing particular traits in response to local environmental conditions (phenotypic plasticity). Organisms can also react to the shifting environment by dispersal; however this option is not always available when, for example, the landscape is too fragmented. The last type of response is evolution via genetic selection leading to adaptation. The persistence of species and populations depends however on the initial response to the shifts in the environment.

The genome of a species contains signatures of these responses that may be studied with genetic markers. We are interested in filling gaps in the knowledge about past demographic history of organisms and the factors that ultimately shape genomic variation in populations. For this we use the very latest, innovative techniques in next generation sequencing. The incorporation of technological developments in molecular biology and the ongoing development of genomic tools, like SNPs and next generation sequencing, and genomic-based approaches, like full genome scans and gene-expression pattern analysis, make it possible to address questions that until now were hard to tackle. There is an urgent need for empirical studies on nonmodel organisms which can contribute to the emerging disciplines of population genomics and landscape genomics. Such studies are necessary if we want to show how these advances in molecular techniques and approaches allow conservation genetics to make a big leap forward. Incorporating genetics and genomics into nature conservation will highlight the following:

(i) Genomic consequences of inbreeding

(ii) Inbreeding by environment interaction

(iii) Genomic and epigenomic consequences of outbreeding

(iv) Genomic and epigenomic mechanisms of phenotypic plasticity

(v) Transcriptome, metabolomics, and proteomic techniques applied to conservation biology

(vi) The emerging discipline of landscape genomics; detection of signature of selection using genomic techniques

All the results from the accepted papers have greatly contributed to the special issue. To optimize the use of limited resources in conservation biology and conservation genomics and minimize the need to sample and disturb wild species (often those most in need of investigation are the most sensitive to disturbance), we recommend increased public sharing of resources including genetic markers, data, and analytical tools and improved professional recognition for the publication of genomic resources.

Acknowledgments

Special thanks also extend to all the authors and referees for their efforts in the special issue.

Cino Pertoldi
Ettore Randi
Aritz Ruiz-González
Philippine Vergeer
Joop Ouborg

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 2152847, 8 pages
http://dx.doi.org/10.1155/2016/2152847

Research Article

Novel Graphical Analyses of Runs of Homozygosity among Species and Livestock Breeds

Laura Iacolina,[1] Astrid V. Stronen,[1] Cino Pertoldi,[1,2] Małgorzata Tokarska,[3] Louise S. Nørgaard,[1] Joaquin Muñoz,[1] Anders Kjærsgaard,[1] Aritz Ruiz-Gonzalez,[4] Stanisław Kamiński,[5] and Deirdre C. Purfield[6]

[1]*Department of Chemistry and Bioscience, Aalborg University, Section of Biology and Environmental Engineering, Fredrik Bajers Vej 7H, 9220 Aalborg, Denmark*
[2]*Aalborg Zoo, Mølleparkvej 63, 9000 Aalborg, Denmark*
[3]*Mammal Research Institute Polish Academy of Sciences, Ul. Waszkiewicza 1, 17-230 Białowieża, Poland*
[4]*Department of Zoology and Animal Cell Biology, University of the Basque Country UPV/EHU, C/Paseo de la Universidad 7, 01006 Vitoria-Gasteiz, Spain*
[5]*Department of Animal Genetics, University of Warmia and Mazury in Olsztyn, 10-718 Olsztyn, Poland*
[6]*Animal & Biosciences Department, Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, County Cork, Ireland*

Correspondence should be addressed to Laura Iacolina; li@bio.aau.dk

Received 14 December 2015; Accepted 28 September 2016

Academic Editor: Graziano Pesole

Runs of homozygosity (ROH), uninterrupted stretches of homozygous genotypes resulting from parents transmitting identical haplotypes to their offspring, have emerged as informative genome-wide estimates of autozygosity (inbreeding). We used genomic profiles based on 698 K single nucleotide polymorphisms (SNPs) from nine breeds of domestic cattle (*Bos taurus*) and the European bison (*Bison bonasus*) to investigate how ROH distributions can be compared within and among species. We focused on two length classes: 0.5–15 Mb to investigate ancient events and >15 Mb to address recent events (approximately three generations). For each length class, we chose a few chromosomes with a high number of ROH, calculated the percentage of times a SNP appeared in a ROH, and plotted the results. We selected areas with distinct patterns including regions where (1) all groups revealed an increase or decrease of ROH, (2) bison differed from cattle, (3) one cattle breed or groups of breeds differed (e.g., dairy versus meat cattle). Examination of these regions in the cattle genome showed genes potentially important for natural and human-induced selection, concerning, for example, meat and milk quality, metabolism, growth, and immune function. The comparative methodology presented here permits visual identification of regions of interest for selection, breeding programs, and conservation.

1. Introduction

Mating among closely related individuals can affect the fitness of the progeny by increasing the inbreeding coefficient (F) [1] and therefore the probability that alleles at a locus, sampled randomly in a population, are identical by descent (IBD) [2]. The reduction in fitness can be due to the accumulation of recessive lethal genetic disorders, reduction of fertility, and lower adaptive potential [1, 3, 4].

In wild living and captive populations, there is an urgent need to reduce inbreeding and augment genetic diversity, and this can be achieved by implementing carefully planned mating strategies. One possibility consists in reducing the level of inbreeding per generation and the response to selection (optimal contribution selection) [5]. The estimation of F requires completeness and accuracy of the available pedigree records, which are not always available, because of missing information or registration errors. When genotypes

are available a probabilistic approach can be utilized for the reconstruction of the pedigree. However, such an approach does not take into account the stochastic nature of recombination [6]. New approaches based on the runs of homozygosity (ROH), which are DNA segments that harbour uninterrupted stretches of homozygous genotypes, have shown to be reliable estimates of autozygosity at the genome-wide level [7–9].

In addition, the frequency and extent of ROH can be used to estimate the time when the inbreeding event took place. Considering that recombination events break long chromosome segments, it is assumed that long autozygous segments in an individual derive from a common recent ancestor, whereas shorter autozygous segments are indicating a remote common ancestor [10–12]. We should therefore expect that the longer the homozygous segments, the more recent the inbreeding. However, long ROH may also be explained by a recent event under strong selective pressure. ROH can thus be used to identify the genomic signatures of recent and/or ancient selective pressure, as shown by [9]. Additionally, fixed ROH in all the individuals in a population could indicate past selective events. Clearly, the presence of long ROH at relatively high frequency in a population could also indicate the presence of genetic substructure, with consanguineous mating occurring only within some subpopulations [13]. ROH are also affected by demographic events [8] and further investigation should examine issues such as skewed reproductive success.

The objective of this study was to describe and compare the distribution of ROH of different length in nine *Bos taurus* cattle breeds under different management practices and selection histories. The same comparison was made at the interspecific level by comparing the distribution of the ROH between the abovementioned cattle breeds and the Lowland line of the European bison (*Bison bonasus*) from the Białowieża National Park (Poland). The Lowland line is highly inbred due to only seven founders [14].

While previous investigations were exclusively based on the count and sum of the number of ROH above a given length [9], in this paper we analysed the frequency of SNPs falling within a ROH above and below an *a priori* chosen length (15 Mb) and we visualized the different distributions across populations. In addition, this graphical visualization allows the identification of similarities and dissimilarities in the regions that can be used to investigate possible adaptive/selective patterns.

2. Material and Methods

2.1. Genotypes and Quality Control. Genotypes consisting of 777,972 single nucleotide polymorphisms (SNPs) from the BovineHD BeadChip (Illumina Inc., San Diego, CA) were generated for 891 sires of multiple breeds. Breeds represented include Angus ($n = 39$), Belgian Blue ($n = 38$), Charolais ($n = 117$), Friesian ($n = 98$), Hereford ($n = 40$), Holstein ($n = 262$), Holstein-Friesian crosses ($n = 111$), Limousin ($n = 128$), and Simmental ($n = 58$) (data from [9]). Angus, Belgian Blue, and Hereford are primarily meat breeds; Friesian, Holstein, and Holstein-Friesian crosses are primarily dairy breeds, while Limousin, Simmental, and Charolais are used for both milk and meat. Forty European Lowland bison (*Bison bonasus*) from Białowieża National Park (Poland) were used for comparison. GenomeStudio™ (Illumina Inc., San Diego, CA) and accompanying guidelines from Illumina (http://www.illumina.com/Documents/products/technotes/technote_infinium_genotyping_data_analysis.pdf) were used for quality control. Total individual call rate in the bison was 0.99. For cattle, only biallelic SNPs on the 29 autosomes were retained after removing all monomorphic SNPs across breeds, filtering for Hardy Weinberg Equilibrium ($p < 0.0001$) within each breed separately and for call rates >90%. Final analyses were performed on 867 cattle and 40 bison with 698,384 SNPs.

2.2. Runs of Homozygosity. Following the approach in [9], ROH were estimated using PLINK v1.07 [15] and were defined within a sliding window of 50 SNPs, in one SNP interval, across the genome. Up to one possible heterozygous genotype was permitted and no more than two SNPs with missing genotypes were allowed per window (see [9]).

ROH were divided in seven length categories (1–5 Mb, 5–10 Mb, 10–15 Mb, 15–20 Mb, 20–25 MB, 25–30 Mb, and >30 Mb). For each ROH length category we summed all ROH per animal and averaged this per cattle breed and for the bison. In order to investigate the potential of our approach, we then focused on two length classes: from 500 Kb till 15 Mb to investigate ancient events and >15 Mb to address recent events. To select target chromosomes for detailed analyses, we created Manhattan plots with SAS 9.4 (SAS Institute Inc., Toronto, Canada) for both length classes and selected the chromosomes accordingly. For the chosen chromosomes, we calculated the percentage of times a SNP appeared in a ROH and plotted these results with SAS.

2.3. Analyses of Genomic Regions in the Runs of Homozygosity. As an example for the methodology applied in this study, we selected regions of the different chromosomes that showed one of the following patterns (see Figure 2): (a) a simultaneous increase (or decrease) in the number of SNPs in a ROH across all populations, as this pattern could possibly involve genes fundamental for the two species analysed; (b) few populations showing an opposite pattern compared to the others, as this could comprise genes specific for those populations; (c) different patterns between dairy and meat breeds, as this could possibly concern regions under human-induced directional selection; (d) different patterns between bison and domestic cattle breeds, as this pattern may be related to traits important for survival in the wild; (e) a single domestic breed differentiating from the others, as this could relate to specific characteristics of that breed; (f) a long region with a high percentage of ROH, as this could be associated with recent selective events; (g) a short region with opposite trend within a longer homogeneous region, to investigate what could have caused such an abrupt change in variability levels. Each region was screened

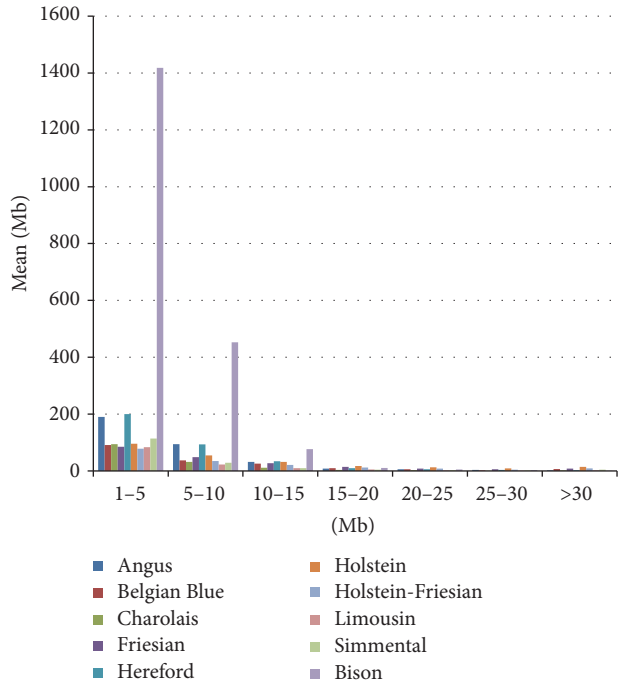


Figure 1: The mean sum of runs of homozygosity (ROH) per genotyped individual, measured in Megabases (Mb) within each population, for each considered ROH length category.

using NCBI (https://www.ncbi.nlm.nih.gov/) resources for the presence of annotated genes (release 104) and information on their biological function and possible evolutionary importance.

3. Results

3.1. Runs of Homozygosity. The European bison exhibited the highest mean sum of ROH in the length categories 1–5 Mb, 5-10 Mb, and 10–15 Mb compared to all the domestic breeds. Angus and Hereford also showed considerably higher mean sums than other breeds in the categories 1–5 Mb and 5–10 Mb (see Figure 1).

In the Manhattan plot for the length class between 500 Kb and 15 Mb, chromosomes 2 and 3 showed a group of extremely variable SNPs, while chromosomes 7, 14, and 16 had the highest density and frequencies of SNPs falling in a ROH (see Figure S1a in Supplementary Material available online at http://dx.doi.org/10.1155/2016/2152847). We thus focused on these chromosomes for subsequent analyses. For the ROH >15 Mb, the Manhattan plot showed a more homogeneous distribution but we selected chromosomes 6, 9, and 20 for subsequent analyses (Figure S2a). In the plots based on ROH < 15 Mb, we observed large regions of the bison genome where almost 100% of SNPs fell within a ROH (Figure S1b–f). The frequency of SNPs falling in a ROH > 15 Mb was lower for all populations, in accordance with the smaller number of ROH in this length category (Figure S2b–d). Additionally, the frequency of a SNP falling within a ROH in the bison was not higher than that observed in the domestic breeds with a single exception on chromosome 9 (Figure S2c). On chromosome 20 the highest percentage of SNPs falling within a ROH was detected in dairy cattle breeds (Figure S2d). No clear pattern was observed on chromosome 6 (Figure S2b).

3.2. Analyses of Genomic Regions in the Runs of Homozygosity. The in-depth analysis of 17 regions, selected from seven chromosomes (i.e., 2, 3, 7, 9, 14, 16, and 20) led to the identification of more than 300 annotated genes whose functions vary considerably (see Table S1). The most frequent functionally characterised genes were those related to metabolic pathways, but we also observed genes related to disease and immune function, growth, and reproduction. As an example, we review here a few of our observations in the selected regions.

In summary, pattern (a) were mainly related to metabolic pathways, involving several CD-, ATP-, and SLAM-family genes (see Table S1) and olfactory receptors. Metabolic pathways were the main genes observed in pattern (b). Pattern (c) was inconclusive for ROH < 15 Mb. In pattern (f) (also an example of (c)) ROH > 15 Mb included genes related to milk and meat quality, growth, and metabolic disorders related to energy unbalanced consumption. Patterns (d) were located in portions of the chromosomes poorly described, with the only exception being the long region on chromosome 9, where a high number of ROH > 15 Mb was observed (Figure S2c). In addition to the metabolism and disease related genes widely encountered in all the screened regions, we report the presence of genes related to olfactory perception, obesity, growth, and sperm malformation in this region. In pattern (e), we observed a region (Figure 2(e)) where the Simmental showed higher variability than the other breeds. Here, genes involved were related to fat thickness and colour, growth, and sperm functionality. In pattern (f), where Hereford showed extremely high frequency values of SNP falling within a ROH and the Belgian Blue extreme variability (with the other breeds in between; Figure S1f, near 45000000), the genes observed were mainly related to the codification of proteins involved in sugar transport and assimilation at cellular level. In pattern (g) we observed genes involved in cortisol pathways and sweet perception, regulation of host response to virus infection, and regulatory function in ovulation.

4. Discussion

Our findings revealed several chromosomes with a high number of ROH, and most results concerned ROH < 15 Mb. Upon closer inspection of selected chromosomes, we observed genes potentially important for natural and human-induced selection, concerning, for example, meat and milk quality, metabolism, growth, and immune function. Hence, the ROH approach appears informative for evaluating and comparing species and population history and evaluating possible patterns of adaptation.

We observed comparatively few results for ROH > 15 Mb, the longer regions that are likely to reflect recent inbreeding

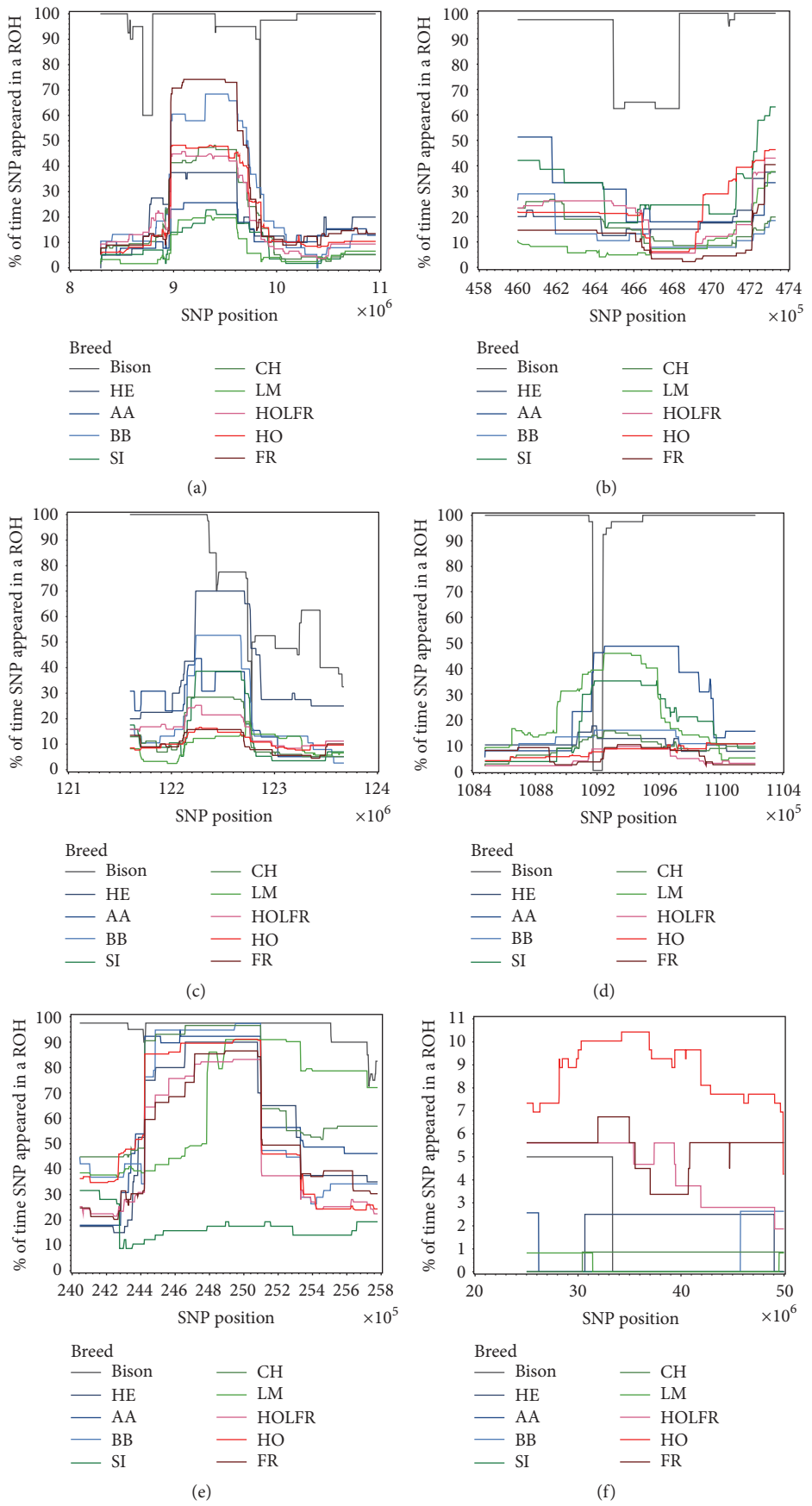


Figure 2: Continued.

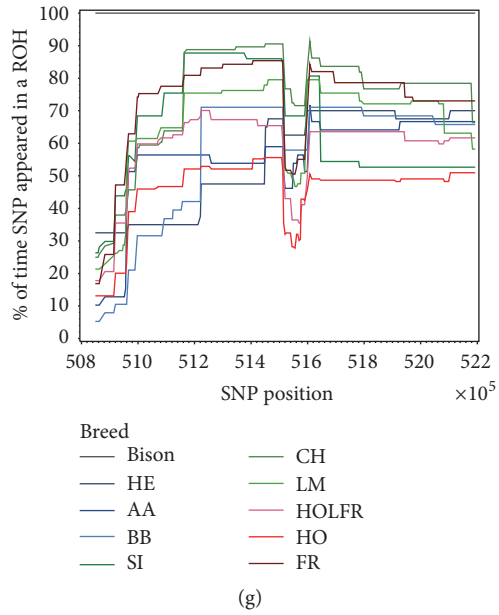


Figure 2: Examples of the investigated patterns. (a) A simultaneous increase (or decrease) in the number of SNPs in a ROH across all populations, as this pattern could possibly involve genes fundamental for the two species analysed (chromosome 3); (b) few populations showing an opposite pattern compared to the others, as this could comprise genes specific for those populations (chromosome 7); (c) different patterns between dairy and meat breeds, as this could possibly concern regions under human-induced directional selection (chromosome 2); (d) different patterns between bison and domestic cattle breeds, as this pattern may be related to traits important for survival in the wild (chromosome 3); (e) a single domestic breed differentiating from the others, as this could relate to specific characteristics of that breed (chromosome 14); (f) a long region with a high percentage of ROH, as this could be associated with recent selective events (chromosome 20); (g) a short region with opposite trend within a longer homogeneous region, to investigate what could have caused such an abrupt change in variability levels (chromosome 7).

[9, 11]. Our results may thus suggest relatively limited recent inbreeding in the cattle breeds included in the study, although the many shorter ROH could indicate a lower N_E in the past [16]. For the European bison, however, large regions of the genome had a 100% (or near 100%) frequency of SNPs falling within a ROH. This suggests high levels of inbreeding, which is consistent with earlier studies and known population history involving a severe bottleneck [17, 18]. However, even limited inbreeding can cause detrimental effects [1, 19] and should be monitored. Earlier studies across species have suggested that ROH > 16 Mb may be considered as recent inbreeding [11, 16]. Analyses of cattle breeds report ROH > 16 Mb as the expected mean after approximately three generations since the most recent common ancestor, whereas autozygosity due to more distant common ancestors will not be captured by this measure [11]. For an in-depth assessment of inbreeding, it may be necessary to investigate different ROH length classes considering the history of the organisms under study. For example, comparisons between wild and domestic species may show different patterns than native and commercial livestock in terms of recent and/or past histories of inbreeding. Consequently, ROH length classes should be assessed on a case by case basis with exploratory analyses informed, where possible, by the history of the species under study.

Variation in sample size and N_E may have influenced the results. Our comparison of, for example, Belgian Blue (n = 38) and Holstein (n = 262) should therefore be interpreted with caution. Other important factors that may play a role are differences in breed genetic diversity. McTavish et al. [20] reported observed heterozygosity for several breeds included in our study based on 50 K SNP markers. Among the breeds that showed distinct ROH patterns in our study, they note that Simmental showed a heterozygosity of 0.28 (n = 77), the Belgian Blue 0.30 (n = 4), the Hereford 0.29 (n = 98), and the Holstein 0.30 (n = 85). Furthermore, the value for Limousin was 0.29 (n = 100) and for Charolais was 0.31 (n = 53). Although these values are similar despite variable sample size, among- and within-breed variation in genetic diversity could affect ROH results and their interpretation and may therefore complicate our comparison of cattle breeds and European bison.

Angus and Hereford breeds, together with bison, show high mean sum of ROH in the length class 1–10 Mb, which may be a result of ancestral relatedness owing to small founder populations and isolated origins [11]. In particular, the ROH for the bison is extremely high for the intervals 1–5 Mb and 5–10 Mb with several regions that are completely fixed. This appears consistent with an estimated N_E of 23 and a total of seven founders for the European bison's Lowland

population [18]. In comparison, a recent survey presented considerably larger but variable census population size (N_C) and N_E for some of the cattle breeds included in our study [21]. For Aberdeen Angus, they reported N_C > 10 M and N_E of 136. For Holstein, N_C was >65 M and N_E was 99, whereas for Limousin, N_C was >4 M and N_E was 174. There may thus be considerable differences in population history among breeds and also for breeds within the same group (such as meat production), which could have affected our results.

We observed genes grouped into various functional categories. The types of genes observed may reveal adaptive patterns and indicate human-induced and/or natural selection, for example, in cases of genes linked to growth and immunity where the first is likely to be human-modified and the second is subject to stronger natural selection. Our results also highlight the need to consider potential conflicts between these two sources of selection. For example, we noted a gene implicated in ketosis (region F, chromosome 20), a metabolic disorder that occurs in cattle when energy demands such as high milk production exceed energy intake and result in a negative energy balance. Strong directional selection for high-performance characteristics such as high milk yield may therefore have implications for animal health and welfare, life expectancy, and the ethical dimensions of animal breeding to cope with their living environments (see, e.g., [22, 23]).

4.1. Applications. The ROH approach seems informative for investigating selection and evolutionary histories across a range of different populations, including wild/domestic species, native/commercial livestock, and commercial breeds of various kinds (e.g., cattle breeds for milk or meat, sheep breeds for meat or wool). Our study compared cattle with one related wild species, the European bison. However, this species is highly inbred and has low genetic diversity [18]. Study of other wild-domestic species pairs may therefore provide a more nuanced picture of genomic regions under selection, for example, in domestic pigs and wild boar, or captive and free-living populations of the wild boar (e.g., [24]), thus taking advantage of recent developments in high-density genomic arrays to investigate domestic and wild species (e.g., [25]).

The results of our analyses may also suggest applications for genetic rescue. This could include key genetic regions of high variability observed in one breed, which could be transferred to one or more other populations, for example, related to immune system function or tolerance to environmental factors such as heat, parasites, and infectious disease [26, 27]. Moreover, genes related to growth may have important applications for animal breeding and could be introduced to new breeds to enhance both genetic variation and production [28]. Further research may also help clarify the extent to which selection for rapid growth might conflict with selection for meat quality, which may be relevant to conservation management and breeding for both commercial and native livestock breeds (e.g., [29]).

It will be important to establish whether ROH are under selection. If a ROH is not under selection, its length should normally decrease with every generation as the expected length of autozygous segments identical by descent follows an exponential distribution with mean equal to $0.5g$ Morgans, where g is the number of generations since the common ancestor [30]. Conversely, a ROH could contain recessive variants that are expressed in the autozygous state. These variants are known to cause various genetic diseases in humans as a result of specific mutations (e.g., phenylketonuria, Tay-Sachs disease, and cystic fibrosis) and may also be involved in complex diseases such as heart and liver diseases and diabetes [31].

For livestock, the incidence of disease associated with intensive production has increased among several breeds [32], such as Holstein and Jersey [33–35]. Additionally, important traits, such as adaptation to low-quality food resources, parasites, and tolerance to disease and temperature fluctuations may be found mostly in native breeds [36]. An important aspect of the ROH assessment will be identification of genetic variants with applications for genetic rescue, which could benefit both native and commercial breeds [28] to increase robustness and tolerance to environmental variation [27, 36].

4.2. Possible Limiting Factors. Ascertainment bias could have affected the comparison of ROH between different species (here cattle and bison) [37]. Moreover, our observations are necessarily incomplete, as there are still large regions of the genome that have not been fully described, as testified by the high number of uncharacterised genes we encountered in our screening (see Table S1). However, key genomic regions can be noted for further research, which also helps identify high-priority areas of the genome for future study.

5. Conclusions

The comparative methodology presented here permits visual identification of regions of interest, which could be of value for selection and breeding programs. The ROH approach offers several immediate applications. Firstly, breeding strategies may be improved by reduction in ROH that are acting to reduce genomic diversity. Such a strategy could be useful where genomic regions have lost important diversity or been accidentally fixed, for example, as a consequence of a population bottleneck and/or founder effect. Further, the ROH approach has implications for genetic rescue and the design of breeding strategies for populations at risk. The presence of ROH at intermediate frequency in a population may indicate heterogeneity of the N_E in different genomic regions. Accordingly, a breeding strategy based on maximising N_E for a population could produce an increase of N_E for some chromosomal regions and a reduction in others. This situation could complicate the design of a long-term protocol because of the risk of fixation of certain genes and loss of genetic diversity. Human-driven breeding could also overwhelm natural selective pressures, especially for populations mainly governed by genetic drift due to the small N_E. It is therefore necessary to balance various considerations for long-term conservation breeding, and information from ROH can help

pinpoint important genomic regions even if we do not, at the moment, have a complete understanding of their function.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

Laura Iacolina and Astrid V. Stronen contributed equally.

Acknowledgments

Cino Pertoldi was supported by a grant from Danish Natural Science Research Council (Grant nos. 11-103926, 09-065999, and 95095995), the Carlsberg Foundation (Grant no. 2011-01-0059), and the Aalborg Zoo Conservation Foundation (AZCF). Laura Iacolina has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Action (Grant Agreement no. 656697). Astrid V. Stronen received funding from the Danish Natural Science Research Council (Postdoctoral Grant 1337-00007).

References

[1] D. W. Bjelland, K. A. Weigel, N. Vukasinovic, and J. D. Nkrumah, "Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding," *Journal of Dairy Science*, vol. 96, no. 7, pp. 4697–4706, 2013.

[2] S. Wright, "Coefficients of inbreeding and relationship," *The American Naturalist*, vol. 56, no. 645, pp. 330–338, 1922.

[3] O. González-Recio, E. López de Maturana, and J. P. Gutiérrez, "Inbreeding depression on female fertility and calving ease in spanish dairy cattle," *Journal of Dairy Science*, vol. 90, no. 12, pp. 5744–5752, 2007.

[4] S. Sommer, "The importance of immune gene variability (MHC) in evolutionary ecology and conservation," *Frontiers in Zoology*, vol. 2, article 16, 2005.

[5] G. Gandini, A. Stella, M. Del Corvo, and G. B. Jansen, "Selection with inbreeding control in simulated young bull schemes for local dairy cattle breeds," *Journal of Dairy Science*, vol. 97, no. 3, pp. 1790–1798, 2014.

[6] R. McQuillan, A.-L. Leutenegger, R. Abdel-Rahman et al., "Runs of homozygosity in European populations," *The American Journal of Human Genetics*, vol. 83, no. 3, pp. 359–372, 2008.

[7] J. Gibson, N. E. Morton, and A. Collins, "Extended tracts of homozygosity in outbred human populations," *Human Molecular Genetics*, vol. 15, no. 5, pp. 789–795, 2006.

[8] M. Bosse, H.-J. Megens, O. Madsen et al., "Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape," *PLoS Genetics*, vol. 8, no. 11, Article ID e1003100, 2012.

[9] D. C. Purfield, D. P. Berry, S. McParland, and D. G. Bradley, "Runs of homozygosity and population history in cattle," *BMC Genetics*, vol. 13, article 70, 2012.

[10] M. Ferencakovic, E. Hamzic, B. Gredler, I. Curik, and J. Sölkner, "Runs of homozygosity reveal genomewide autozygosity in the Austrian Fleckvieh cattle," *Agriculturae Conspectus Scientificus*, vol. 76, no. 4, pp. 325–328, 2011.

[11] M. Ferenčaković, E. Hamzić, B. Gredler et al., "Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations," *Journal of Animal Breeding and Genetics*, vol. 130, no. 4, pp. 286–293, 2013.

[12] M. Ferenčaković, J. Sölkner, and I. Curik, "Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors," *Genetics Selection Evolution*, vol. 45, no. 1, article 42, 2013.

[13] L. Iacolina, M. Scandura, D. J. Goedbloed et al., "Genomic diversity and differentiation of a managed island wild boar population," *Heredity*, vol. 116, no. 1, pp. 60–67, 2016.

[14] H. M. Slatis, "An analysis of inbreeding in the European bison," *Genetics*, vol. 45, pp. 275–287, 1960.

[15] S. Purcell, B. Neale, K. Todd-Brown et al., "Plink: a tool set for whole-genome association and population-based linkage analyses," *American Journal of Human Genetics*, vol. 81, no. 3, pp. 559–575, 2007.

[16] M. Kirin, R. McQuillan, C. S. Franklin, H. Campbell, P. M. McKeigue, and J. F. Wilson, "Genomic runs of homozygosity record population history and consanguinity," *PLoS ONE*, vol. 5, no. 11, Article ID e13996, 2010.

[17] K. Oleński, M. Tokarska, D. M. Hering et al., "Genome-wide association study for *posthitis* in the free-living population of European bison (*Bison bonasus*)," *Biology Direct*, vol. 10, no. 2, 2015.

[18] M. Tokarska, C. Pertoldi, R. Kowalczyk, and K. Perzanowski, "Genetic status of the European bison *Bison bonasus* after extinction in the wild and subsequent recovery," *Mammal Review*, vol. 41, no. 2, pp. 151–162, 2011.

[19] J. E. Pryce, M. Haile-Mariam, M. E. Goddard, and B. J. Hayes, "Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle," *Genetics Selection Evolution*, vol. 46, article 71, 2014.

[20] E. J. McTavish, J. E. Decker, R. D. Schnabel, J. F. Taylor, and D. M. Hillis, "New World cattle show ancestry from multiple independent domestication events," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 15, pp. E1398–E1406, 2013.

[21] The Bovine HapMap Consortium, "Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds," *Science*, vol. 324, no. 5926, pp. 528–532, 2009.

[22] W. M. Muir, H.-W. Cheng, and C. Croney, "Methods to address poultry robustness and welfare issues through breeding and associated ethical considerations," *Frontiers in Genetics*, vol. 5, article 407, 2014.

[23] W. M. Rauw and L. Gomez-Raya, "Genotype by environment interaction and breeding for robustness in livestock," *Frontiers in Genetics*, vol. 6, article 310, 2015.

[24] A. Canu, S. Costa, L. Iacolina, P. Piatti, M. Apollonio, and M. Scandura, "Are captive wild boar more introgressed than free-ranging wild boar? Two case studies in Italy," *European Journal of Wildlife Research*, vol. 60, no. 3, pp. 459–467, 2014.

[25] V. R. Kharzinova, A. A. Sermyagin, E. A. Gladyr, I. M. Okhlopkov, G. Brem, and N. A. Zinovieva, "A study of applicability of SNP chips developed for bovine and ovine species to whole-genome analysis of reindeer *Rangifer tarandus*," *Journal of Heredity*, vol. 106, no. 6, pp. 758–761, 2015.

[26] F. Alpay, Y. Zare, M. H. Kamalludin et al., "Genome-wide association study of susceptibility to infection by *Mycobacterium avium* subspecies *paratuberculosis* in Holstein cattle," *PLoS ONE*, vol. 9, no. 12, Article ID e111704, 2014.

[27] L. R. Porto-Neto, A. Reverter, K. C. Prayaga et al., "The genetic architecture of climatic adaptation of tropical cattle," *PLoS ONE*, vol. 9, no. 11, Article ID e113284, 2014.

[28] T. N. Kristensen, A. A. Hoffmann, C. Pertoldi, and A. V. Stronen, "What can livestock breeders learn from conservation genetics and *vice versa*?" *Frontiers in Genetics*, vol. 5, article 38, 2015.

[29] G. Holló, K. Nuernberg, T. Somogyi, I. Anton, and I. Holló, "Comparison of fattening performance and slaughter value of local Hungarian cattle breeds to international breeds," *Archiv für Tierzucht*, vol. 55, no. 1, pp. 1–12, 2012.

[30] D. P. Howrigan, M. A. Simonson, and M. C. Keller, "Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms," *BMC Genomics*, vol. 12, article 460, 2011.

[31] L. B. Andrews, J. E. Fullarton, N. A. Holtzman, and A. G. Motulsky, *Assessing Genetic Risks: Implications for Health and Social Policy*, National Academy Press, Washington, DC, USA, 1994.

[32] I. Medugorac, C. E. Veit-Kensch, J. Ramljak et al., "Conservation priorities of genetic diversity in domesticated metapopulations: a study in taurine cattle breeds," *Ecology and Evolution*, vol. 1, no. 3, pp. 408–420, 2011.

[33] J. S. Agerholm, C. Bendixen, O. Andersen, and J. Arnbjerg, "Complex vertebral malformation in Holstein calves," *Journal of Veterinary Diagnostic Investigation*, vol. 13, no. 4, pp. 283–289, 2001.

[34] P. M. VanRaden, K. M. Olson, D. J. Null, and J. L. Hutchison, "Harmful recessive effects on fertility detected by absence of homozygous haplotypes," *Journal of Dairy Science*, vol. 94, no. 12, pp. 6153–6161, 2011.

[35] C. Charlier, J. S. Agerholm, W. Coppieters et al., "A deletion in the bovine FANCI gene compromises fertility by causing fetal death and brachyspina," *PLoS ONE*, vol. 7, no. 8, Article ID e43085, 2012.

[36] I. Hoffmann, "Climate change and the characterization, breeding and conservation of animal genetic resources," *Animal Genetics*, vol. 41, no. 1, pp. 32–46, 2010.

[37] L. K. Matukumalli, C. T. Lawley, R. D. Schnabel et al., "Development and characterization of a high density SNP genotyping assay for cattle," *PLoS ONE*, vol. 4, no. 4, Article ID e5350, 2009.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 3654093, 10 pages
http://dx.doi.org/10.1155/2016/3654093

Research Article

Using Genome-Wide SNP Discovery and Genotyping to Reveal the Main Source of Population Differentiation in *Nothofagus dombeyi* (Mirb.) Oerst. in Chile

Rodrigo Hasbún,[1] Jorge González,[1] Carolina Iturra,[1] Glenda Fuentes,[2] Diego Alarcón,[2,3] and Eduardo Ruiz[2]

[1]*Departamento de Silvicultura, Facultad de Ciencias Forestales, Universidad de Concepción, 4070386 Concepción, Chile*
[2]*Departamento de Botánica, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, 4070386 Concepción, Chile*
[3]*Instituto de Ecología y Biodiversidad (IEB), Facultad de Ciencias, Universidad de Chile, 7800003 Santiago, Chile*

Correspondence should be addressed to Rodrigo Hasbún; rodrigohasbun@udec.cl

Received 29 January 2016; Accepted 23 May 2016

Academic Editor: Ettore Randi

Within a woody plant species, environmental heterogeneity has the potential to influence the distribution of genetic variation among populations through several evolutionary processes. In some species, a relationship between environmental characteristics and the distribution of genotypes can be detected, showing the importance of natural selection as the main source of differentiation. *Nothofagus dombeyi* (Mirb.) Oerst. (Nothofagaceae) is an endemic tree species occurring both in Chile and in Argentina temperate forests. Postglacial history has been studied with chloroplast DNA and evolutionary forces shaping genetic variation patterns have been analysed with isozymes but fine-scale genetic diversity studies are needed. The study of demographic and selection histories in *Nothofagus dombeyi* requires more informative markers such as single nucleotide polymorphisms (SNP). Genotyping-by-Sequencing tools now allow studying thousands of SNP markers at reasonable prices in nonmodel species. We investigated more than 10 K SNP loci for signatures of local adaptation and showed that interrogation of genomic resources can identify shifts in genetic diversity and putative adaptive signals in this nonmodel woody species.

1. Introduction

In population genetics and conservation the big question is genetic drift instead of natural selection. Definitely both processes determine evolution, but genetic drift operates randomly and depends on effective population size while natural selection proceeds nonrandomly and relies on environmental variables. The evolution towards hereditary adaptations to the current environment is determined by natural selection and has a direction; genetic drift instead is governed solely by chance. Consequently, drift acts on alleles, which generally have no phenotypic effect; instead selection favours certain alleles that increase fitness, reduce the unfavourable alleles frequencies, and ignore neutral alleles [1]. Knowing what the main microevolutionary force is is very relevant for rational genetic management of threatened species, especially for species with geographical distribution severely fragmented [2].

Approaches to addressing adaptive variation have been incorporated into the definition of evolutionary significant units [3, 4]. New technologies like Next Generation Sequencing (NGS) and fine-scale GIS, coupled with advances in computer hardware and software in the field of genomics [5–7], have allowed the development of new methods for comprehensive evaluation of adaptive diversity [8, 9]. The discovery and genotyping of massive genetic markers are now enabled by modern genomic tools at very low cost. This makes the study of adaptive genetic loci possible on a wide range of species, which can facilitate the identification of key biodiversity areas. Kirk and Freeland [10] reviewed

some of the applications of neutral versus adaptive markers in molecular ecology, discussed some of the advantages that can be obtained by supplementing studies of molecular ecology with data from nonneutral molecular markers, and summarized new methods that allow generating data from loci under selection. Population genomic analyses require multilocus datasets from multiple populations and identification of nonneutral or adaptive loci by contrasting patterns of population divergence among genetic regions.

Studies in nonmodel organisms have shown relatively broad candidate genomic regions that are under selection, but it remains difficult to identify the genes (or the mutations) that are affected by selection. Increasing the density of markers in genome scans is paramount to overcome this problem, and validating signals of selection from particular genes using multiple methods should also help [4]. One of the most exciting developments in population genomics is the development of various reduced-representation protocols, collectively referred to as Genotyping-by-Sequencing (GBS), which allow sequencing of a subset of the genome through selective amplification of restriction fragments [10].

Nothofagus dombeyi (Mirb.) Oerst. (Nothofagaceae) is an endemic tree species occurring both in Chile and in Argentina temperate forests with a remarkably broad altitudinal and latitudinal distribution, across many different ecological gradients in the former [11]. The evergreen tree *N. dombeyi* is a pioneer species and constitutes an important element in the dynamics of South American forest. Its postglacial history has been studied with chloroplast DNA and evolutionary forces shaping genetic variation patterns have been analysed with isozymes [12, 13]. However, genome-wide scan methods using thousands of markers to study a representative portion of the genome are needed.

In this work, we assess GBS in the nonmodel woody species *N. dombeyi* to develop high quantity of informative markers such as single nucleotide polymorphisms (SNP). Our aim is to determine the contribution of selection and molecular adaptation to shaping genome-wide variation. We expect higher genetic population differentiation (for ecological separated localities) for adaptive SNP than neutral SNP if natural selection is the principal source of differentiation. Alternatively, if other sources of differentiation (mutation, genetic drift, and migration) are relevant, they will equally affect both types of SNP. Knowing the contribution of selection effect to shaping genome variation patterns will have many applications for biodiversity conservation, especially in endangered species, because neutral and adaptive genetic diversity will likely have different impacts on long-term survival. In fact, in most cases, only adaptive diversity will allow a population to adapt to changing environmental conditions.

2. Materials and Methods

2.1. Sampling Design

2.1.1. Niche Modelling. To consider the remarkably broad altitudinal and latitudinal distribution of this species, across many different ecological gradients, we used the method proposed by Alarcón and Cavieres [14] to niche modelling of *Nothofagus dombeyi* in Chile using BIOMOD [15]. Eight variables were selected from the WorldClim global climate database [16] corresponding to the present climate conditions with a 30-arc-second grid resolution, with the least correlation among them for the studied species range area. Four variables were related to energy constraints: (a) BIO2: mean diurnal temperature range; (b) BIO4: temperature seasonality; (c) BIO5: maximum temperature in the warmest month; and (d) BIO6: minimum temperature in the coldest month. Other four variables were related to water availability: (e) BIO12: annual precipitation; (f) BIO15: precipitation seasonality; (g) BIO18: precipitation in the warmest quarter; and (h) BIO19: precipitation in the coldest quarter. Then, we projected the current and future distribution (year 2050) considering a conservative future climate projection CSIRO B2A 2050 by using the tools of BIOMOD software. Further, we identified geographical areas with potential habitat loss, which should be identified as high priority in genetic conservation programs.

2.1.2. Ecological Regions or Strata. Relatively homogeneous units in ecological terms (strata) were defined from natural populations of the species associated with the geographical areas projected in its ecological niche modelling. The Calinski-Harabasz criterion, which is a pseudo-F statistic as in ANOVA, was used to assess the best number of strata identified by K-means partitioning [17].

2.2. DNA Sample and Library Preparation

2.2.1. Plant Material and Genomic DNA Isolation. Adult trees with a diameter higher than 50 cm were sampled during the growing season of 2013-2014 from twenty-one sites covering almost the entire range of the species *N. dombeyi* in Chile. One to four sites were assigned to each stratum according to its superficies and 2 to 9 samples per site were taken (Table 1; Supplementary Figure 1 in Supplementary Material available online at http://dx.doi.org/10.1155/2016/3654093).

DNA extraction was performed using a Qiagen DNeasy Plant kit (Qiagen Inc., USA). Lyophilized leaf tissue (20 mg) was ground in a Precellys®24 (Precellys, USA) homogenizer with two 1/4″ ceramic spheres (MP BIOMEDICALS, USA) and AP1 buffer. The objective is grinding tissues and lysing cells prior to DNA extraction. DNA extraction protocol was done following the manufacturer instructions but elution was done with 30 μL (instead of 100 μL) to increase the final DNA concentration in the eluate (>100 ng/μL). The integrity of genomic DNA was evaluated by agarose gel and quantified using a Qubit fluorometer (Invitrogen, USA).

2.2.2. Library Preparation and High-Throughput Sequencing. Library preparation and high-throughput sequencing were performed at University of Wisconsin Biotechnology Center (DNA Sequencing Facility). The GBS genomic library preparation was done following the protocol detailed by Elshire et al. [18] with the methylation-sensitive restriction

Table 1: Strata code and location, geographical coordinates, and sample size of the sampled individuals of *Nothofagus dombeyi* in Chile.

Stratum	Location (code)	Latitude	Longitude	*N*
1	Altos Lircay (AL)	−35.599162	−71.044414	4
1	Antuco (AN)	−37.343457	−71.615626	5
2	Ralco (RA)	−37.925041	−71.575168	6
2	Termas de Tolhuaca (TT)	−38.235047	−71.727552	9
3	Lago La Paloma (LP)	−45.876213	−72.070813	4
3	El Machi (EM)	−45.009553	−71.906872	5
3	Near Villa Amengual (NVAM)	−45.008498	−71.908366	2
3	Villa Amengual (VAM)	−45.007328	−71.911175	5
4	Nonguén (NO)	−36.879745	−72.987923	5
4	Villa Las Araucarias (VA)	−36.879774	−72.987981	6
4	Caramávida (CA)	−36.879774	−72.987981	4
5	Mariquina (MA)	−39.471811	−73.055799	4
5	Lago Riñihue (LRI)	−39.480193	−73.048809	3
5	Fundo Llancahue (FL)	−39.858686	−73.141572	3
5	Lago Neltume (LN)	−39.859282	−73.141650	3
5	Parque Nacional Villarrica (PNV)	−39.341073	−71.972351	5
5	Loncoche Interior (LI)	−39.341820	−71.972232	3
5	Melipeuco (ME)	−38.912949	−71.704088	3
6	Las Trancas (LT)	−40.221101	−73.362430	5
6	Camino Osorno a Maicolpué (COM)	−40.598321	−73.497015	4
7	Parque Nacional Puyehue (PNP)	−40.737218	−72.306050	3
7	Lago Rupanco (LR)	−40.736785	−72.302922	5

enzyme ApeKI and 96 custom barcodes. Illumina high-throughput sequencing was conducted on an Illumina HiSeq 2000 (Illumina, USA) using 100 bp single-end sequencing runs. The samples were sequenced across one Illumina lane. Base calling was performed in Casava v1.8.2 (Illumina, USA).

2.3. Nonreference SNP Calling

2.3.1. De Novo Identification of Loci/Alleles. Sequence results were analysed and SNP genotypes were assigned using the UNEAK (Universal Network Enabled Analysis Kit) GBS pipeline [19], which is part of the TASSEL 3.0 [20] bioinformatic analysis package. This pipeline does not depend on a reference sequence, which is the actual case for *N. dombeyi*. SNP discovery is performed directly within pairs of matched sequence tags (unique sequence representing a group of reads) and filtered through network analysis. The network filter trimmed reads to 64 bp to reduce the effects of error sequencing and enabled efficient storage of data in bit format. SNP were assigned with default settings. Briefly, tags differing by a single nucleotide were retained as SNP and those with a minor allele frequency 0.05 were removed to minimize the impact of sequencing errors [19]. We used a minimum call rate of 0 and additional filters were applied in next steps.

2.3.2. Post-SNP Calling Filters and Imputation. Given that we are trying to find SNP for population genetic analysis, we applied some filters to remove loci and individuals that contain very low levels of information prior to further analysis. We applied two functions of TASSEL 5.2 that removed all SNP (rows) and then samples (columns) containing 90% or more "*N*" values (indicating that neither allele is designated). These *N*s represent individuals where the allele cannot be called from the sequence reads. This is because either no read is available at this site (for this individual) or the sequence quality is too low to be called.

In order to cope with missing data, genotype imputation was used to fill in the missing data and improve the power of downstream analyses. We used LinkImpute implemented in TASSEL 5.2, a software package based on a *k*-nearest neighbour genotype imputation method, LD-*k*NNi [21]. This imputation method was designed specifically for nonmodel organisms in which genomic resources are poorly developed and marker order is unreliable or unknown.

2.4. Detection of Selection Footprints. To identify adaptive SNP (putative loci under selection), we used LOSITAN [22]. LOSITAN is a selection detection workbench based on the Fst-outlier methods. We used 50,000 simulations, 0.99 for confidence interval, false discovery rate of 0.05, mutation model "Infinite Alleles," and the options "Neutral mean Fst" and "Force mean Fst", which iteratively identify and remove Fst outliers when calculating the global distribution of Fst. Our interest is in patterns of adaptation driven by environmental gradients; therefore we focused on outlier patterns

indicating divergent selection (Fst significantly higher than neutral expectations).

LOSITAN analyses were complemented with BayeScan 2.1 [23] for estimating the posterior probability that a given locus is affected by selection. Briefly, prior odds of 100 were used for identifying the top candidates of the selected loci and a total of 50,000 reversible-jump Markov Chain Monte Carlo chains were run with a thinning interval of 10, following 20 pilot runs of 5,000 iterations each, and a burn-in length of 50,000. Loci were considered outliers with an FDR of 0.05.

To confirm the adaptive SNP detected by previous methods, the spatial analysis method (SAM) implemented in the program Samβada v0.5.1 [24] was used. We conducted the analysis using the 10,109 SNP detected for *N. dombeyi* in Chile. Samβada uses logistic regressions to model the probability of presence of an allelic variant in a polymorphic marker given the environmental conditions of the sampling locations. Eight environmental variables previously described in Section 2.1.1 were used (temperature related → BIO2, BIO4, BIO5, and BIO6; precipitation related → BIO12, BIO15, BIO18, and BIO19). Regarding genotypes, each of the states of a given SNP is considered independently as binary presence/absence in each sample. Our biallelic SNP were recoded as three distinct genotypes (AA, AB, and BB). A maximum likelihood approach is used to fit the models using univariate analyses. Each model for a given genotype is compared to a constant model, where the probability of presence of the genotype is the same at each location. The statistical significance threshold was set to 1% before applying Bonferroni correction. Significance was assessed with log likelihood ratio (G) tests [25] selecting loci/allele that tested higher than the 99th percentile of the G score distribution.

2.5. Estimation of Genome-Wide Genetic Variation and Differentiation. We made all estimations in parallel with neutral SNP (10,109) and adaptive SNP.

2.5.1. Basic Statistics of Genetic Variation. All the results were obtained using the adegenet [26, 27] and hierFstat [28] R packages. Basic statistics were estimated including observed heterozygosity (Ho) and genetic diversity (Hs) within population. Also, overall gene diversity (Ht) and corrected Ht (Htp), gene diversity among samples (Dst), and corrected Dst (Dstp) were estimated. Fst and corrected Fst (Fstp) were assessed as well as Fis following Nei [29] per overall loci. Dest, a measure of population differentiation as defined by Jost [30], was also calculated. The degree of genetic differentiation among populations is expected to be low for neutral SNP but highly divergent in SNP subject to directional selection.

2.5.2. Population Structure. To describe the genetic biodiversity of a species, more important than diversity among individuals is the diversity between groups of individuals. First we analysed individual data to identify populations, or more large genetic clusters, and then we described these clusters by adegenet R package. To get a simplified picture of the genetic diversity observed among individuals or populations we used Principal Component Analysis (PCA). Discriminant Analysis of Principal Components (DAPC) function was used to describe the relationships between these clusters. The main results of DAPC were DAPC scatterplots.

For each pair of strata, we computed pairwise Fst values with hierFstat R package. Principal Coordinates Analysis (PCoA) on Fst values was performed to detect major genetic clusters (K) at individual level.

2.5.3. Detecting Locus Contributions. In DAPC, the variables actually analysed are principal components of a PCA. Loadings of these variables are generally uninformative, since PCs themselves do not all have straightforward interpretations. However, we can also compute contributions of the alleles, which can turn out to be very informative. In general, there are many alleles and their contribution is best plotted for a single discriminant function at a time using adegenet R package.

3. Results and Discussion

3.1. Niche Modelling and Strata Definition. Modelled area of *N. dombeyi* presence under present climate conditions is 88,174 km^2 and in the future (year 2050) under dispersal constraints is 72,928 km^2. Although it is not qualified with a status of threatened species, we estimated a decrease of almost 20% of its habitat area in a relatively short time, particularly in the northern populations, associated with Mediterranean-influenced climate, which is at least worrisome considering we used the most conservative future climate scenario. Fuss et al. [31] showed that the actual reality is the worst climate scenario projected. The K-means analyses, applied to four best least correlated variables of WorldClim (BIO2, BIO4, BIO12, and BIO15), identified seven strata groups of *N. dombeyi* in Chile at very broad spatial scales (Figure 1). These strata represented relatively homogeneous habitats according to the climatic key variables mentioned. The strata named Septentrional (#1), Alto Biobío (#2), Los Lagos Andes (#7), and Patagonia (#3) form the latitudinal gradient across Andes Mountain range including higher altitudes, while the strata Nahuelbuta (#4), Araucanía y Los Ríos (#5), and Los Lagos Costa (#6) present some oceanic influence at a gradient located in Coastal Mountain range and central valley of the central distribution of *N. dombeyi* at lower elevation. The strata were coherently identified as the bioclimatic transition from temperate biome with some Mediterranean influence, especially in the northern portion of the Septentrional stratum and then the typical temperate bioclimatic classification across the rest of the distribution of the species [32].

3.2. Assessment of N. dombeyi Genotypes Sampled by GBS Using ApeKI. We obtained more than 17 Gb out of 172,171,356 reads, from which 85.82% keep a score $Q \geq 30$. From the 96 barcoded samples in our study, an average of 1.5 million (SD 1.7 million) sequence reads per sample were obtained (range from 278 to 8.2 millions) (see Supplementary Figure 2). After removal of SNP with a minor allele frequency 0.05 and removing samples and SNP with more than 90% "NN" (unassigned) genotypes, the dataset consisted of 73

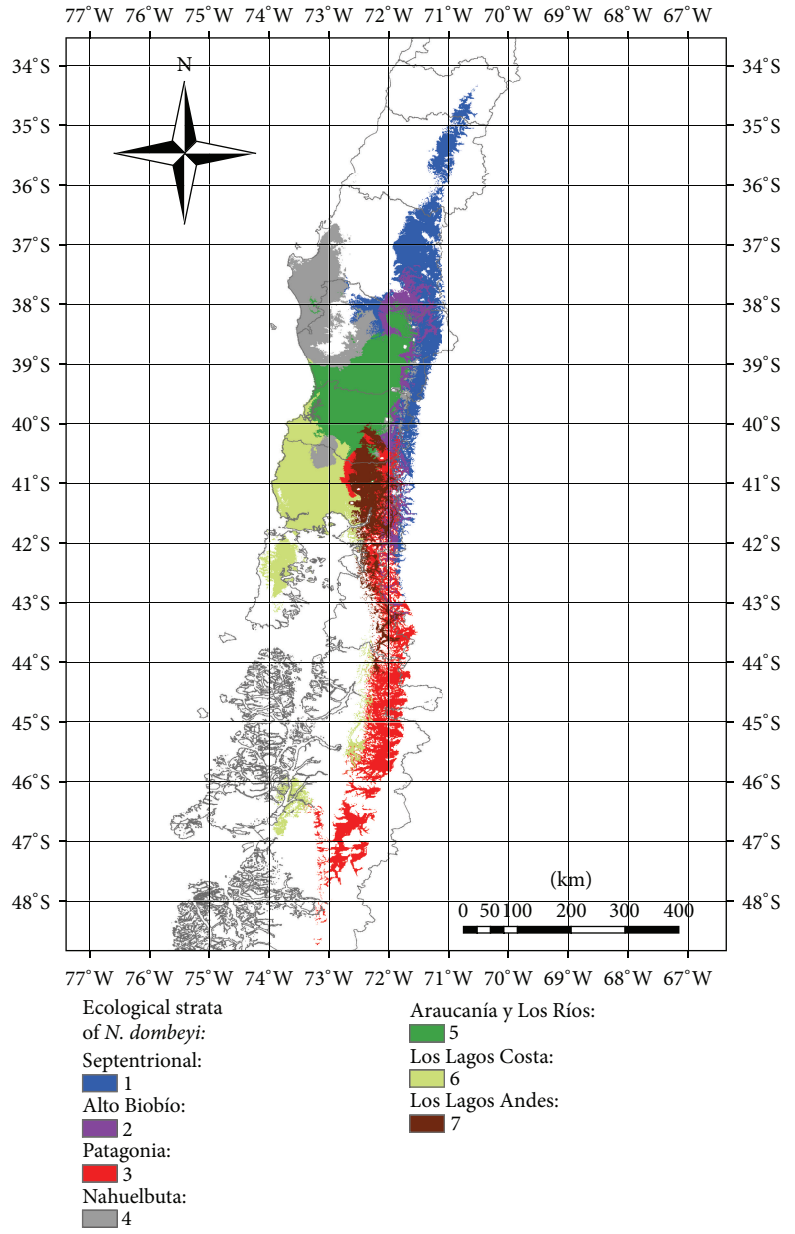


Figure 1: Using ecological niche models to identify different strata for *Nothofagus dombeyi* in Chile.

individuals with 10,109 binary SNP from seven strata. This dataset was subjected to the impute step and finally total numbers of "NN" genotypes were 19,067 (0.03%) missing data.

3.3. Putative under Selection SNP (Adaptive SNP). Using the LOSITAN approach, 124 adaptive SNP (1.2%) were identified (directionally selected) at the false discovery rate (FDR) threshold of 0.1 (Table 2; Supplementary Table 1). We detected 99 adaptive SNP in BayeScan and there were 48 overlap cases between the loci reported by each method (see Supplementary Table 1). The adaptive SNP detected by both approaches were limited (39%), possibly reflecting discrepancies in their methodologies. However, all the 124 SNP identified by LOSITAN had Fst values ≥ 0.30 (estimated across the 73 *N. dombeyi* individuals and 10,109 loci), which may be considered as strongly differentiated.

Using Samβada we detected 2,406 significant genotypes associated with a given environmental variable. They represent 884 SNP (8.7% of total SNP assessed) and only 3% correspond to heterozygous genotypes. From the 124 outlier SNP identified by LOSITAN, genotypes in 121 SNP were identified as associated with environmental variables by Samβada (see Supplementary Table 1). Genetic differentiation associated with both temperature and precipitation gradients was detected.

Table 2: Representative list of outlier single nucleotide polymorphisms (SNP) as putative candidates for adaptation in *Nothofagus dombeyi* in Chile and their significant associations with environmental variables using Samβada. Italic values show an SNP that is not a significant outlier.

Locus	SNP ID	LOSITAN			BayeScan	Samβada							
		Het	Fst	P (Simul Fst < sample Fst)		BIO2	BIO4	BIO5	BIO6	BIO12	BIO15	BIO18	BIO19
TP3479	4	0.368	0.545	0.99999	Yes	x		x			x		x
TP8429	14	0.282	0.496	0.99996	—	x		x			x		x
TP23119	—	0.491	0.035	*0.46921*									
TP30254	42	0.385	0.442	0.99996	Yes	x		x			x		x
TP30852	43	0.438	0.438	1.00000	Yes	x		x			x		x
TP32296	44	0.467	0.377	0.99991	Yes	x		x			x	x	
TP38613	50	0.364	0.618	1.00000	Yes	x		x			x		x
TP65805	91	0.540	0.488	1.00000	Yes		x		x				

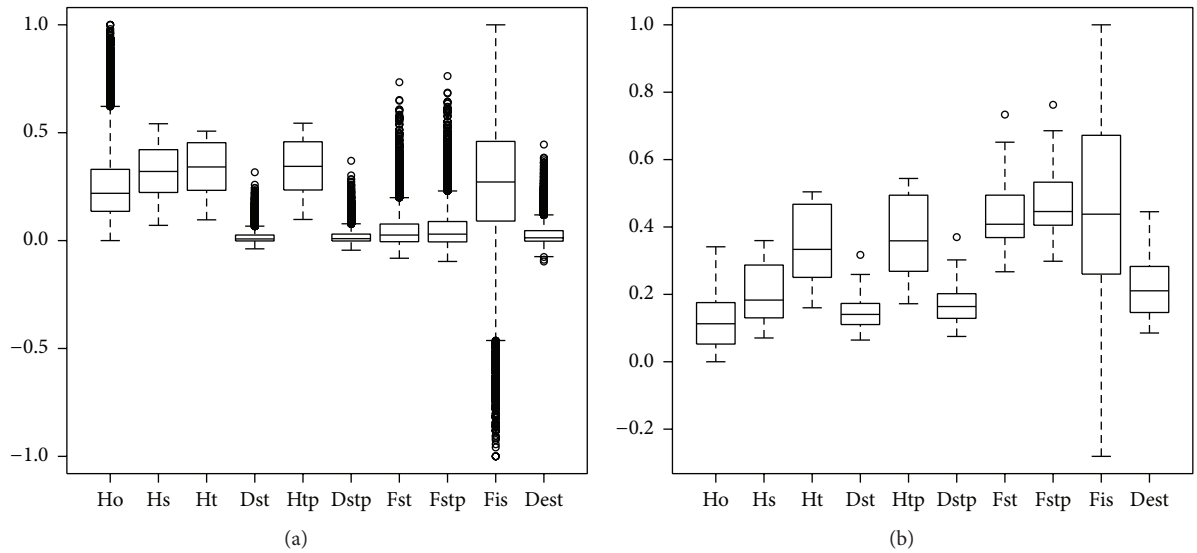


Figure 2: Summary statistics of genetic variation existing in *Nothofagus dombeyi* in Chile estimated by 10,109 neutral SNP (a) or 124 adaptive SNP (b). Ho: heterozygosity within population; Hs: genetic diversity within population; Ht: overall gene diversity; Http: corrected Ht; Dst: gene diversity among samples; Dstp: corrected Dst; Fst: fixation index; Fstp: corrected Fst; Fis: inbreeding coefficient per overall loci; Dest: measure of population differentiation.

3.4. Genome-Wide Genetic Variation and Differentiation. Using neutral plus outlier SNP (10,109), different basic statistics of genetic variation of *N. dombeyi* in Chile show low to medium genetic diversity level and low level of strata differentiation (Figure 2). The low genetic structure found indicates relatively high gene flow, which is consistent with the fact that *N. dombeyi* has a distribution more or less continuous and is wind pollinated. This result is similar to other results obtained with the same and other species of *Nothofagus* gender [33, 34]. With outlier SNP, the indicators based on heterozygosity increase and F statistics shows clear differences within strata.

Overall average pairwise Fst values among populations, calculated over the set of the 10,109 SNP, show medium population structuring across the distribution of *N. dombeyi* in Chile (Table 3). The average pairwise Fst values ranged from 0.062 between two very close strata (4 and 5) in the Coastal Mountains to 0.258 between strata 1 and 3, both in Andean Mountains but in the extremes of distribution (Table 1).

Table 3: Average pairwise Fst among seven strata with *Nothofagus dombeyi* presence in Chile based on a set of 10,109 single nucleotide polymorphisms (SNP). Fst 0.15–0.2 are moderately differentiated in bold font and Fst > 0.25 are considered strongly differentiated in bold and italic font.

Strata	1	2	3	4	5	6
2	0.074					
3	***0.258***	**0.193**				
4	0.111	0.074	**0.183**			
5	0.109	0.066	0.138	0.062		
6	**0.156**	0.108	0.140	0.096	0.069	
7	**0.172**	0.121	0.141	0.110	0.081	0.096

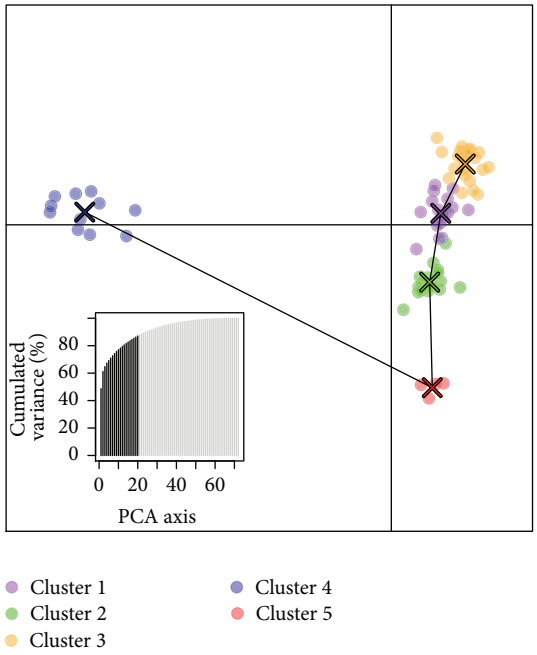


Figure 3: Discriminant Analysis of Principal Components (DAPC) scatterplot drawn using 124 outlier single nucleotide polymorphisms (SNP) across 73 *Nothofagus dombeyi* individuals in the R package adegenet. Dots represent individuals, with colours denoting cluster allocation. Percentages of cumulated variance explained by Principal Component 1 (PC1) to PC20 are shown in the bottom left corner. Minimum spanning tree based on the (squared) distances between clusters within the entire space is shown.

The Principal Coordinates Analysis (PCoA) plot revealed a geographically ordered pattern (see Supplementary Figure 3). The first PC suggests the existence of two clades in the data, while the second one shows groups of closely related isolates arranged along a cline of genetic differentiation. Premoli [35] found continuous clinal genetic variation in populations of *N. pumilio* along the altitudinal gradient, as a result of adaptive responses to ecological gradients and/or restrictions for gene flow. The same pattern is shown by neutral SNP and adaptive SNP but with a more clear delimitation in the latter case. This structure was confirmed by a neighbour-joining (NJ) tree (Supplementary Figure 4). Again higher resolution is achieved with adaptive SNP. As expected, both approaches give congruent results, but both are complementary; NJ shows bunches of related individuals, but the cline of genetic differentiation is much clearer in PCA.

Visualisation of broad-scale population structure using a DAPC (Figure 3 and Table 4) with 124 outlier SNP revealed two distinct genetic groups. One group included individuals sampled in all strata corresponding to Clusters 1, 2, 3, and 5 (Table 4), and the other group comprised only ten individuals of stratum 3 plus one individual of stratum 2 forming Cluster 4 (represented almost exclusively by individuals of Patagonia stratum). Also the actual proximities between clusters show the great genetic distance of Cluster 4 with the rest. This result is consistent with a north–south phylogenetic divergence at

Table 4: Individual allocation to five genetic clusters according to strata determined by niche modelling of *N. dombeyi* in Chile. Clusters were identified by find.cluster function of adegenet R package using 124 adaptive single nucleotide polymorphisms (SNP).

Strata	Cluster assignation				
	C1	C2	C3	C4	C5
1	—	—	8	—	—
2	—	—	12	1	—
3	—	—	—	10	4
4	8	—	1	—	—
5	10	5	1	—	—
6	—	7	—	—	—
7	—	6	—	—	—

c. 43°S found within subgenus *Nothofagus* in southern South America, including *N. dombeyi* [12, 36].

Clearly, when examining patterns of *N. dombeyi* population differentiation at neutral versus adaptive SNP, we have detected several distinct differences. Primarily stochastic processes drive differentiation of neutral SNP, whereas both selective and stochastic processes drive that of adaptive SNP, for example, [37–39]. This result supports the hypothesis that putative under selection SNP are being affected by adaptation, which is not affecting the neutral SNP. The next step is to characterize the phenotypic outcomes of alternative genotypes to confirm if the mechanism is genetic drift or selection, similar to previous studies in *Nothofagus pumilio* [40, 41]. Ideally, phenotype-genotype association studies should be followed by the profiling of gene expression, functional tests, and selection tests to determine that a gene or genes are involved in shaping an adaptive trait [42].

3.5. Loci Contributions. Over 24% of the adaptive SNP detected by LOSITAN had a high loading (here defined as ≥0.02) in one or two of the three PC in a PCA with more than 10 K loci of the Chilean population of *N. dombeyi* (Supplementary Figures 5, 6, and 7), thus confirming its contribution to population structure. Adaptive SNP ID 4, 14, 42, 50, and 91 are five of the most discriminating loci and their allele frequencies over seven strata show a high or even dominant allele frequency in stratum 3 (Figure 4). This is the southernmost stratum (#3 Patagonia) and this may justify the differences in allele frequencies. Irrespective of the mechanism underlying these changes (drift or selection), this illustrates that, in the natural distribution of *N. dombeyi* in Chile, specific nucleotides can undergo drastic changes within only a hundred kilometres of distance. Our results exhibited spatial patterns differentiating one stratum in loci which could occur in genomic regions or genes for important functions of *N. dombeyi* across their habitat. For example, the GBS sequence tags of adaptive SNP ID 43 (TP30852) show a high identity with a heat shock protein or adaptive SNP ID 44 (TP32296) with histone acetyltransferase enzyme (data not shown). The first is fundamental for heat or other environmental stresses and the second can be involved in drought sensing or another response to environmental

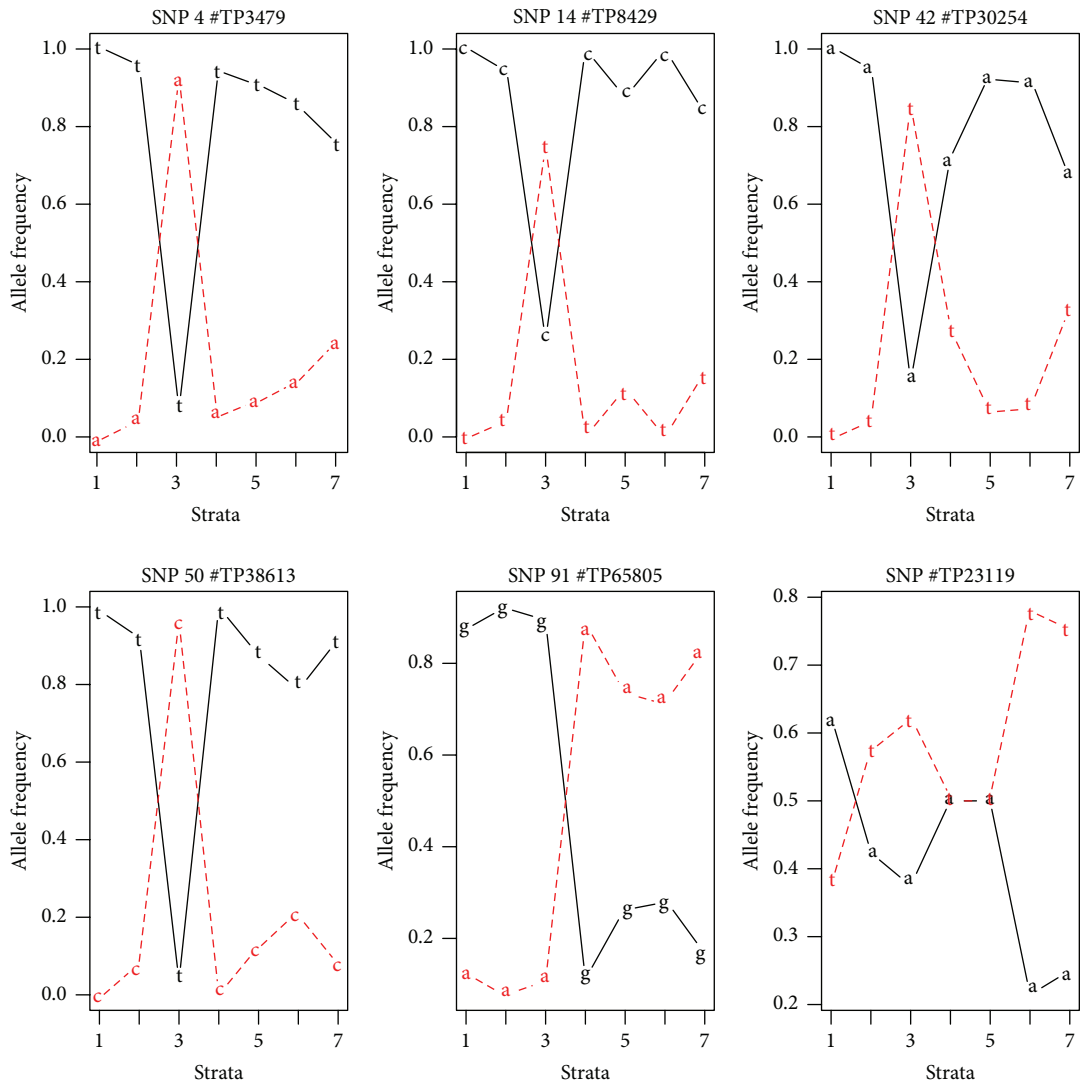


FIGURE 4: Spatial distributions of single nucleotide polymorphism (SNP) loci/genotypes in *N. dombeyi* in Chile for each strata. The graphs show frequencies of loci/genotypes differentiating among strata (SNP #3479, SNP #8429, SNP #30254, SNP #38613, and SNP #65805) and no differentiating among strata (SNP #23119).

factors. We therefore suggest that these two SNP may be interesting candidates for future functional studies that might facilitate other studies, for example, [43] focused on the development of genomic resources for *Nothofagus* species. However, improvements in study design and analyses of replicated studies will be needed before this very promising approach can be brought to application for managing genetic resources [44].

4. Conclusions

Development of 10,109 genome-wide SNP for *N. dombeyi* using GBS made evaluation of genomic diversity and fine-scale population structure possible for the first time in this species in Chile. Results showed that genome-wide patterns of genetic diversity and differentiation varied widely across the genome. As such, we identified numerous genomic regions exhibiting signatures of divergent selection. We have also provided strong evidence of substantial genetic differentiation associated with both temperature and precipitation gradients, suggesting local adaptation.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

The authors thank CONAF (Corporación Nacional Forestal) and SNASPE (Sistema Nacional de Áreas Silvestres Protegidas del Estado) for providing permission to sample plant material and especially thank José Luis Pérez for

sampling stratum 3 (Patagonia). Angela Sierra, Daniela Fernandez, Giannina Espinoza, and Joseline Valdés are gratefully acknowledged for providing technical and scientific assistance. The authors thank the University of Wisconsin Biotechnology Center DNA Sequencing Facility for providing sequencing and support services. Funding for this project was provided by the Research Fund for Native Forest of CONAF (Project 068/2012).

References

[1] R. Frankham, J. D. Ballou, and D. A. Briscoe, *A Primer of Conservation Genetics*, Cambridge University Press, Cambridge, UK, 2004.

[2] R. Frankham, "Where are we in conservation genetics and where do we need to go?" *Conservation Genetics*, vol. 11, no. 2, pp. 661–663, 2010.

[3] D. J. Fraser and L. Bernatchez, "Adaptive evolutionary conservation: towards a unified concept for defining conservation units," *Molecular Ecology*, vol. 10, no. 12, pp. 2741–2752, 2001.

[4] B. Gebremedhin, G. F. Ficetola, S. Naderi et al., "Frontiers in identifying conservation units: from neutral markers to adaptive genetic variation," *Animal Conservation*, vol. 12, no. 2, pp. 107–109, 2009.

[5] F. W. Allendorf, P. A. Hohenlohe, and G. Luikart, "Genomics and the future of conservation genetics," *Nature Reviews Genetics*, vol. 11, no. 10, pp. 697–709, 2010.

[6] C. Y. Tang, C.-L. Hung, H. Zheng, C.-Y. Lin, and H. Jiang, "Novel computational technologies for next-generation sequencing data analysis and their applications," *International Journal of Genomics*, vol. 2015, Article ID 254685, 3 pages, 2015.

[7] B. B. Patnaik, S. Y. Park, S. W. Kang et al., "Transcriptome profile of the asian giant hornet (*Vespa mandarinia*) using illumina HiSeq 4000 sequencing: *De novo* assembly, functional annotation, and discovery of SSR markers," *International Journal of Genomics*, vol. 2016, Article ID 4169587, 15 pages, 2016.

[8] A. Bonin, F. Nicole, F. Pompanon, C. Miaud, and P. Taberlet, "Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation," *Conservation Biology*, vol. 21, no. 3, pp. 697–708, 2007.

[9] D. Demontis, C. Pertoldi, V. Loeschcke, K. Mikkelsen, T. Axelsson, and T. N. Kristensen, "Efficiency of selection, as measured by single nucleotide polymorphism variation, is dependent on inbreeding rate in *Drosophila melanogaster*," *Molecular Ecology*, vol. 18, no. 22, pp. 4551–4563, 2009.

[10] H. Kirk and J. R. Freeland, "Applications and implications of neutral versus non-neutral markers in molecular ecology," *International Journal of Molecular Sciences*, vol. 12, no. 6, pp. 3966–3988, 2011.

[11] A. C. Premoli, M. P. Quiroga, P. Mathiasen, and T. Kitzberger, "Ecological niche modeling meets phylogeography to unravel hidden past history of key forest genera in plant geography: *Podocarpus* and *Nothofagus*," *Natureza & Conservação*, vol. 10, no. 2, pp. 160–168, 2012.

[12] M. C. Acosta, P. Mathiasen, and A. C. Premoli, "Retracing the evolutionary history of Nothofagus in its geo-climatic context: new developments in the emerging field of phylogeology," *Geobiology*, vol. 12, no. 6, pp. 497–510, 2014.

[13] A. C. Premoli, P. Mathiasen, M. Cristina Acosta, and V. A. Ramos, "Phylogeographically concordant chloroplast DNA divergence in sympatric Nothofagus s.s. How deep can it be?" *New Phytologist*, vol. 193, no. 1, pp. 261–275, 2012.

[14] D. Alarcón and L. A. Cavieres, "In the right place at the right time: habitat representation in protected areas of South American *Nothofagus*-dominated plants after a dispersal constrained climate change scenario," *PLoS ONE*, vol. 10, no. 3, Article ID e0119952, 2015.

[15] W. Thuiller, B. Lafourcade, R. Engler, and M. B. Araújo, "BIOMOD—a platform for ensemble forecasting of species distributions," *Ecography*, vol. 32, no. 3, pp. 369–373, 2009.

[16] R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis, "Very high resolution interpolated climate surfaces for global land areas," *International Journal of Climatology*, vol. 25, no. 15, pp. 1965–1978, 2005.

[17] T. Calinski and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, pp. 1–27, 1974.

[18] R. J. Elshire, J. C. Glaubitz, Q. Sun et al., "A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species," *PLoS ONE*, vol. 6, no. 5, Article ID e19379, 2011.

[19] F. Lu, A. E. Lipka, J. Glaubitz et al., "Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol," *PLoS Genetics*, vol. 9, no. 1, Article ID e1003215, 2013.

[20] J. C. Glaubitz, T. M. Casstevens, F. Lu et al., "TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline," *PLoS ONE*, vol. 9, no. 2, article e90346, 2014.

[21] D. Money, K. Gardner, Z. Migicovsky, H. Schwaninger, G.-Y. Zhong, and S. Myles, "LinkImpute: fast and accurate genotype imputation for nonmodel organisms," *G3: Genes, Genomes, Genetics*, vol. 5, no. 11, pp. 2383–2390, 2015.

[22] T. Antao, A. Lopes, R. J. Lopes, A. Beja-Pereira, and G. Luikart, "LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method," *BMC Bioinformatics*, vol. 9, article 323, 2008.

[23] M. Foll and O. Gaggiotti, "A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective," *Genetics*, vol. 180, no. 2, pp. 977–993, 2008.

[24] S. Joost, A. Bonin, M. W. Bruford et al., "A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation," *Molecular Ecology*, vol. 16, no. 18, pp. 3955–3969, 2007.

[25] A. J. Dobson and A. G. Barnett, *An Introduction to Generalized Linear Models*, Chapman and Hall/CRC, London, UK, 3rd edition, 2008.

[26] T. Jombart, "adegenet: a R package for the multivariate analysis of genetic markers," *Bioinformatics*, vol. 24, no. 11, pp. 1403–1405, 2008.

[27] T. Jombart and I. Ahmed, "adegenet 1.3-1: new tools for the analysis of genome-wide SNP data," *Bioinformatics*, vol. 27, no. 21, Article ID btr521, pp. 3070–3071, 2011.

[28] J. Goudet, "HIERFSTAT, a package for R to compute and test hierarchical F-statistics," *Molecular Ecology Notes*, vol. 5, no. 1, pp. 184–186, 2005.

[29] M. Nei, *Molecular Evolutionary Genetics*, Columbia University Press, New York, NY, USA, 1987.

[30] L. Jost, "GST and its relatives do not measure differentiation," *Molecular Ecology*, vol. 17, no. 18, pp. 4015–4026, 2008.

[31] S. Fuss, J. G. Canadell, G. P. Peters et al., "Betting on negative emissions," *Nature Climate Change*, vol. 4, no. 10, pp. 850–853, 2014.

[32] F. Luebert and P. Pliscoff, *Sinopsis Bioclimática y Vegetacional de Chile*, Editorial Universitaria, Santiago, Chile, 2006.

[33] A. C. Premoli, "Genetic variation in a geographically restricted and two widespread species of South American *Nothofagus*," *Journal of Biogeography*, vol. 24, no. 6, pp. 883–892, 1997.

[34] C. J. Duncan, J. R. P. Worth, G. J. Jordan, R. C. Jones, and R. E. Vaillancourt, "Genetic differentiation in spite of high gene flow in the dominant rainforest tree of southeastern Australia, *Nothofagus cunninghamii*," *Heredity*, vol. 116, no. 1, pp. 99–106, 2016.

[35] A. C. Premoli, "Isozyme polymorphisms provide evidence of clinal variation with elevation in *Nothofagus pumilio*," *Journal of Heredity*, vol. 94, no. 3, pp. 218–226, 2003.

[36] A. C. Premoli, P. Mathiasen, M. C. Acosta, and V. A. Ramos, "Phylogeographically concordant chloroplast DNA divergence in sympatric *Nothofagus s.s.* How deep can it be?" *New Phytologist*, vol. 193, no. 1, pp. 261–275, 2012.

[37] J. Galindo, P. Morán, and E. Rolán-Alvarez, "Comparing geographical genetic differentiation between candidate and non-candidate loci for adaptation strengthens support for parallel ecological divergence in the marine snail *Littorina saxatilis*," *Molecular Ecology*, vol. 18, no. 5, pp. 919–930, 2009.

[38] A. V. Stronen, B. Jedrzejewska, C. Pertoldi et al., "Genome-wide analyses suggest parallel selection for universal traits may eclipse local environmental selection in a highly mobile carnivore," *Ecology and Evolution*, vol. 5, no. 19, pp. 4410–4425, 2015.

[39] B. Guo, J. DeFaveri, G. Sotelo, A. Nair, and J. Merilä, "Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks," *BMC Biology*, vol. 13, article 19, 2015.

[40] A. C. Premoli and C. A. Brewer, "Environmental v. genetically driven variation in ecophysiological traits of *Nothofagus pumilio* from contrasting elevations," *Australian Journal of Botany*, vol. 55, no. 6, pp. 585–591, 2007.

[41] A. C. Premoli, E. Raffaele, and P. Mathiasen, "Morphological and phenological differences in *Nothofagus pumilio* from contrasting elevations: evidence from a common garden," *Austral Ecology*, vol. 32, no. 5, pp. 515–523, 2007.

[42] C. Pardo-Diaz, C. Salazar, and C. D. Jiggins, "Towards the identification of the loci of adaptive evolution," *Methods in Ecology and Evolution*, vol. 6, no. 4, pp. 445–464, 2015.

[43] V. A. El Mujtar, L. A. Gallo, T. Lang, and P. Garnier-Géré, "Development of genomic resources for *Nothofagus* species using next-generation sequencing data," *Molecular Ecology Resources*, vol. 14, no. 6, pp. 1281–1295, 2014.

[44] I. Ćalić, F. Bussotti, P. J. Martínez-García, and D. B. Neale, "Recent landscape genomics studies in forest trees—what can we believe?" *Tree Genetics & Genomes*, vol. 12, no. 1, pp. 1–7, 2016.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 2374610, 12 pages
http://dx.doi.org/10.1155/2016/2374610

Review Article

Integrating Genomic Data Sets for Knowledge Discovery: An Informed Approach to Management of Captive Endangered Species

Kristopher J. L. Irizarry,[1,2] Doug Bryant,[3] Jordan Kalish,[1] Curtis Eng,[1] Peggy L. Schmidt,[1] Gini Barrett,[1] and Margaret C. Barr[1]

[1]*College of Veterinary Medicine, Western University of Health Sciences, 309 East Second Street, Pomona, CA 91766, USA*
[2]*The Applied Genomics Center, Graduate College of Biomedical Sciences, Western University of Health Sciences, 309 East Second Street, Pomona, CA 91766, USA*
[3]*Danforth Center, 975 N. Warson Road, St. Louis, MO 63132, USA*

Correspondence should be addressed to Kristopher J. L. Irizarry; kirizarry@westernu.edu

Received 24 December 2015; Revised 19 March 2016; Accepted 21 March 2016

Academic Editor: Ettore Randi

Many endangered captive populations exhibit reduced genetic diversity resulting in health issues that impact reproductive fitness and quality of life. Numerous cost effective genomic sequencing and genotyping technologies provide unparalleled opportunity for incorporating genomics knowledge in management of endangered species. Genomic data, such as sequence data, transcriptome data, and genotyping data, provide critical information about a captive population that, when leveraged correctly, can be utilized to maximize population genetic variation while simultaneously reducing unintended introduction or propagation of undesirable phenotypes. Current approaches aimed at managing endangered captive populations utilize species survival plans (SSPs) that rely upon mean kinship estimates to maximize genetic diversity while simultaneously avoiding artificial selection in the breeding program. However, as genomic resources increase for each endangered species, the potential knowledge available for management also increases. Unlike model organisms in which considerable scientific resources are used to experimentally validate genotype-phenotype relationships, endangered species typically lack the necessary sample sizes and economic resources required for such studies. Even so, in the absence of experimentally verified genetic discoveries, genomics data still provides value. In fact, bioinformatics and comparative genomics approaches offer mechanisms for translating these raw genomics data sets into integrated knowledge that enable an informed approach to endangered species management.

1. Introduction

Today's technology makes it feasible to sequence the genome of almost any species of interest and to investigate complex genetic relationships in populations of animals [1–3]. However, the task of translating enormous amounts of genetic data into practical applications is still a work in progress [4–6]. As the ability to predict the functional impact of genetic variations within a population improves, genomics data could provide a powerful tool for informing the management of captive endangered species.

Whole genome sequencing has the potential to be very effective in conservation efforts [2]. In the past, genetic studies have focused on a few loci from many individuals; genomics now allows the focus to switch to many genes from a few individuals. This is particularly important when studying endangered species and working with very limited sample sizes. In addition, genomics provides us with the opportunity to look at the entire genome and not just the pieces that directly code for proteins in attempts to better understand if they contribute to survival. Overall, genomics has the opportunity to play a pivotal role in proactively understanding pressures

and potential stressors that are leading some species into extinction [7].

Endangered species can benefit greatly from the use of comparative genomics. Using the genome sequence from close relatives as reference, the genome of endangered species can be compiled with relatively few samples, which is particularly important when only a limited number of animals exist in captivity. Once the genome is sequenced, identifying deleterious mutations caused by single nucleotide variants (SNVs) provides a valuable resource for management of captive populations [8]. For example, the California condor diverged from the chicken over 100 million years ago but comparison of the genomes reveals useful information in the management of this endangered species. A BAC library was created for the condor and comparative genomics was employed, identifying 93 genes that were aligned with the chicken chromosome map [9]. Using this approach, several genes involved in bone and cartilage formation have been identified, and further tests may reveal the role these genes play in the condor chondrodystrophy, a heritable embryonic lethal phenotype that is present in the population. Understanding the genetic basis of this disease greatly increases the ability to manage and progress the care of this captive population [10]. Another finding in these birds is a mutation in one of the estrogen receptors that has been linked to altered receptor activation by endocrine-disrupting chemicals, such as dichlorodiphenyltrichloroethane (DDT) and its metabolites. This mutation may affect their sensitivity to such chemicals, resulting in reproductive disorders [11]. A comparative genomics approach could be used to identify additional genes associated with reproduction that might also be impacted by environmental contaminants and other factors, providing a multifaceted view of reproductive issues in these animals.

Major histocompatibility complex (MHC) genes are the most polymorphic genes in the vertebrate genome and can be helpful markers in identifying population diversity as well as indirectly measuring immunologic fitness. Ujvari and Belov [12] emphasize the use of markers like MHC to better plan conservation and manage captive populations. For example, MHC has been implicated in mate choice and pregnancy outcomes, which could be very important when working with captive breeding programs. The MHC class II B polymorphism was investigated in a number of wild populations of gorilla via noninvasive fecal sample collection coupled with next-generation sequencing [13]. The analysis led to the identification of 18 different alleles that had not been previously characterized in the gorilla populations. MHC genes have been identified as markers for immune function and survival in almost all vertebrates, and genetic variation of this region can lead to increased or decreased disease susceptibilities [14, 15].

Genome sequencing has enabled huge progress in understanding disease even in very understudied subjects. The Tasmanian devil has recently suffered a major population crash from the result of devil facial tumor disease, and genomics approaches have been implemented to further understand the specifics of the disease and potentials for more effective treatments [16]. Hawaiian honeycreepers are an adaptive radiation that contains numerous endangered species and is of particular interest due to their diverse phenotypes and heterogeneous responses to avian malaria. Genome sequencing and assembly with SNP discovery are providing insight into why some of these birds are malaria-resistant and others are not [17].

In attempts to preserve some of the most endangered species, zoos and other organizations have established captive breeding programs and in some cases reintroduced endangered species back into the wild. Genomics has the potential to facilitate assessing the genetic fitness of individuals within a population and assist programs in correctly identifying the most successful breeding pairs to ensure genetic diversity among future generations. Identifying the best founder individuals in a captive breeding program can greatly increase the success of the program and potentially enhance fitness of the species [18–21].

2. Types of Genomics Data and Analyses

A variety of genomics data types exist. Genomic sequence data provides the foundation of an annotation framework anchored to primary nucleotide sequence organized around contigs and scaffolds comprising individual chromosomes. Sets of sequence intervals corresponding to genes, exons, introns, promoters, enhancers, untranslated regions (UTRs), and intergenic regions comprise the organizational structure of genomics sequence data. Protein coding gene level annotation is structured around the one-to-many relationship that potentially exists between genomic loci and their paired RNA transcripts and protein sequences. Together, the set of transcripts derived from a single genomic locus represents the sequence complexity of gene expression. Additional levels of complexity arise through the tissue and cell specific patterns of expression associated with each transcript as well as the combination of cis regulatory elements responsible for both the transcript sequence and expression patterns. Transcriptome sequencing can provide a wealth of information about the pattern of gene expression including the alternative isoforms, protein coding single nucleotide polymorphisms (SNPs), and tissue specific patterns of expression important in development, health, and disease [22].

Genetic variation is an integral component of genomics data and represents the ability to investigate relationships that may exist between genes, tissues, individuals, and phenotypes [23]. Among the genetic variants that exist are repetitive elements, copy number variations, single nucleotide substitutions, single nucleotide insertions/deletions (indels), genomic inversions, and genomic duplications, to name a few. The most prevalent variants are SNPs, and once a sizable set has been identified, relationships between allelic variation and phenotypes can be empirically investigated.

Traditionally, genotyping in conservation genetics was accomplished using individual assays targeting a single polymorphism, such as restriction fragment length polymorphism (RFLP) analysis [24]. As the number of polymorphisms assayed increased, PCR-based methods of genotyping replaced the more cumbersome RFLP methods. In parallel,

repeat polymorphisms can also be interrogated via PCR-based sequencing methods to take advantage of the increased value associated with multiallelic markers. As the number of polymorphisms under investigation approaches thousands to hundreds of thousands, nucleotide hybridization chips become the platform of choice [25]. Other methods, such as genotyping by sequencing (GBS) [26] and SNP analysis via transcriptome sequencing [27], offer opportunities not only to genotype known polymorphisms, but also to detect de novo and private variants that are unique to a specific individual. Sequencing methods of polymorphism discovery have the potential to identify hundreds of thousands to millions of SNPs, depending upon the depth of sequencing coverage and the relative extent of genetic diversity represented in the samples [28].

Over the past decade, genomic data acquisition has become increasingly routine. Since the year 2005, data output from next-generation sequencing (NGS) platforms has more than doubled each year [25]. Whole human genomes worth of sequence data can be produced as cheaply as $1000, a staggering decrease from the original human genome project's cost of nearly $3 billion, and produced over days or hours instead of what was originally many years [29]. NGS technology has enabled a multitude of scientific investigations that previously could never have even been considered.

Genomics analyses in general, including comparative genomics studies, depend heavily on having an available high-quality genome reference sequence. Unfortunately, unlike model organisms such as mouse, endangered species typically lack significant or any available genomic resources [3]. An important first step in genomics-enabled endangered species management is, therefore, development of a high-quality genome reference sequence. While it is not an insignificant undertaking, reference genome sequence development is becoming more and more commonplace.

Reference genome sequence development starts with raw sequence data (Figure 1). While reference sequences of large complex eukaryotic species constructed entirely from single molecule sequencing technology is becoming more common [30], typically such projects are largely based on paired-end and mate-pair data from NGS platforms. Paired-end data results from sequencing the paired-ends of inserts in the range of hundreds of base pairs, while mate-pair data results from sequencing the ends of inserts in the range of thousands of base pairs. Sufficient coverage from both library types is imperative for the resulting genomic reference sequence to be of sufficient quality. These genome sequence data are then processed and assembled [31] using appropriate methods based on actual data types, with the resulting assembly constituting the species draft genome sequence.

Once a draft genome sequence is constructed, structural annotation can proceed (Figure 1). Here, features such as genomic repeats, genes, and noncoding RNAs are identified, a process that can be informed by closely related model organisms. For example, the well-annotated mouse and human genomes and structural annotations can be used to help identify genes in mammalian endangered species, resulting in higher quality gene and other feature calls than what would be available without such additional information. Genomic

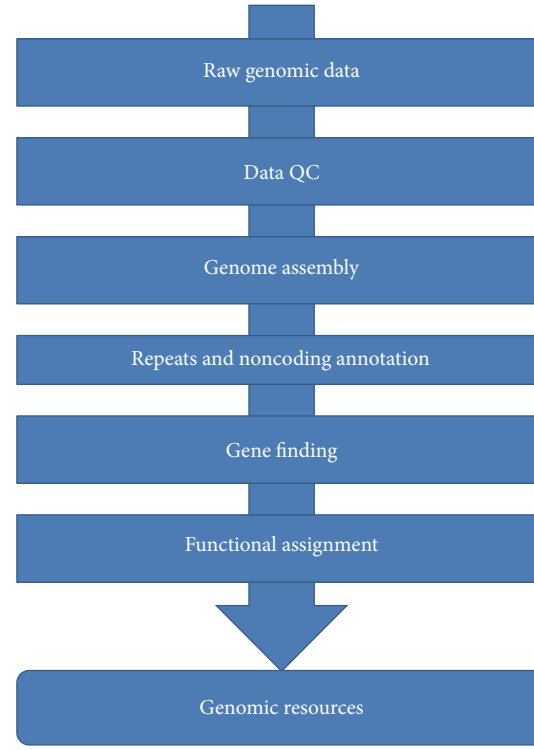


Figure 1: Overview of genomic resource development steps. Raw data is quality controlled and then assembled. Repeats and noncoding features are structurally annotated and used to mask the genome sequence. Genes are then called on this masked sequence, followed by functional assignment. Genome assembly, repeats, noncoding features, and genes constitute basic genomic resources.

data from more closely related species (e.g., domestic bovine genome for endangered bovine species) can be used, if available [32]. Typically, repeats are first identified and used to mask the draft genome sequence, a step necessary to help reduce false positive gene calls resulting from transposable elements and other nongene features. Noncoding RNA molecules are also identified, generally with homology- and motif-based methods. Genes are then identified on the repeat masked genome sequence, informed by closely related model organismal sequence, and, if transcriptome data from the target endangered species is available, these transcript data as well. Once genes have been identified, functional assignments can be made via homology to related species functional annotation, homology versus compiled databases such as the NCBI NR database (http://www.ncbi.nlm.nih.gov/refseq/), the Uniprot database (http://www.uniprot.org/), Conserved Elements from Genomic Alignments (CEGA, http://cega.ezlab.org/) [33], and motif-based methods such as those leveraged with InterProScan (http://www.ebi.ac.uk/Tools/pfa/iprscan5/).

These genomic resources, including the (1) draft genome sequence, (2) structural annotation, and (3) functional annotation, form the basis of all subsequent genomic analyses (Figure 1). Whereas a high depth of sequencing is required in development of the draft genome sequence itself, comparatively very low coverage is all that is needed to assess genomic

traits of further individuals. While Illumina recommends 30x coverage for accurate single nucleotide polymorphism (SNP) and short insertion and deletion (indel) identification, it is typical to perform such analyses using much lower coverage on the order of 10–15x. It is common to sequence tens of individuals together (provided with identification labels such as multiplex identifiers, MDs) in a single sequencing unit and then bioinformatically separate by individual and analyze the resulting sequence data on a per-individual basis [34]. Such methods form a powerful mechanism on which we analyze large numbers of individuals for very low cost. A per-individual and practically exhaustive genomic fingerprint, or the identification of all SNPs and indels in each individual, allows direct comparison between all available members in an endangered species population [35–37].

3. Comparative Genomics Approaches

Just as comparative physiology and comparative anatomy offer a context for appreciating the mechanisms underlying variation in form and function, comparative genomics aids in elucidating conserved and divergent genetic mechanisms associated with specific phenotypes. Comparative genomics methods represent undervalued, yet extremely powerful tools for exploring patterns of shared and divergent biology between pairs of organisms' genomes [38]. Unlike well-studied model organisms, such as mouse and dog, captive endangered species are relatively poorly studied, and the possibility of developing knockouts or transgenic strains of endangered species expressly for experimental exploration of their biology is not a feasible option. Subsequently, the rate at which functionally important genomic signals are identified in endangered species lags significantly behind the rate for typical model organisms. In fact, sometimes genomics sequence data provides relatively little benefit for endangered species conservation efforts, even well after the genomics resources have been produced and deposited into a public database/repository. This is not a consequence of poor genomic quality, but rather the challenge in translating the raw genomic data into management informing knowledge.

Of particular interest are comparative genomics approaches that can rapidly identify functionally important genomic regions with implications for health and disease. The mouse is one of the most widely studied genetic models in the world, resulting in the production of mouse lines having mutations in over one-third of the genes encoded in the mouse genome [39]. The International Mouse Phenotyping Consortium (IMPC) is a collaborative functional genomics effort between laboratories in America, Germany, United Kingdom, France, Canada, China, and Japan. The IMPC has characterized phenotype data for approximately 2000 mouse genes and plans to have a total of 5000 genes characterized by 2016 [40]. The Mouse Genome Database (MGD) is a central repository for mouse functional genomics data and resources including phenotype annotations for mouse genes. Functional annotation in the form of ontologies, such as the Gene Ontology and Mammalian Phenotype Ontology [41], is integrated with the mouse genome.

The tremendous wealth of mouse genomic data can be employed to enable discovery in endangered species through ortholog-based mapping (Figure 2) of mouse phenotype annotations onto endangered species genomes (Figure 3). Such an approach would provide a set of one-to-one orthologs in an endangered species with which phenotypes experimentally identified in the mouse can be associated. Combining these phenotype associated orthologs with functional genetic variation, such as missense mutations in critical residues of highly conserved domains and nonsense or frameshift mutations occurring at the N-terminal portion of protein coding genes, offers high confidence candidates for alleles likely to modulate specific phenotypes. These potential genotype-phenotype relationships can serve as the foundation for identifying members of the endangered species population that may be at risk for undesirable phenotypes.

Genomes from domesticated species, such as the dog, cat, and chicken, have also proved useful in the management of endangered species. For example, comparative genomics approaches using domestic cat MHC loci have been used to quantify MHC diversity in endangered felids including the Florida panther and the cheetah. This information provides a metric for assessing population susceptibility to emerging immunological threats such as bacteria and viruses [42]. Evolutionary conservation of short tandem repeat polymorphisms between domestic cat and cheetah facilitated the creation of PCR-primers capable of amplifying polymorphic loci in the cheetah to determine population-level genetic diversity in this endangered species [43].

Similarly, the dog genome has been used to identify short tandem repeat markers in the maned wolf, a threatened species in Brazil [44]. The red wolf is an endangered species that suffers from coyote introgression. Microsatellite markers have been developed to identify hybrids in the population and remove them in an effort to conserve the red wolf [45].

The value of model genomes for management of endangered species was exemplified in the development of chicken-condor comparative physical maps [9, 10] that were subsequently used to ultimately develop genetic tests for identifying condors at risk for producing offspring with undesirable phenotypes. The utility of comparative genomics in the conservation of endangered species will continue to be valuable, especially as new and improved tools and resources for comparative genomics are developed in the future.

4. Enhanced Management via Collaborative Genetic Association Studies

The identification of genetic variants within genes implicated in specific phenotypes provides a framework for identifying members of the endangered captive population that might be at risk for clinically relevant phenotypes [46]. It is important to make the distinction between validated genetic associations that are identified in typical genetic association studies and the bioinformatics based identification of genotype-phenotype candidates [47]. These candidates are not proven to be associated with the phenotypes. However, the comparative genomics association these orthologs have with their

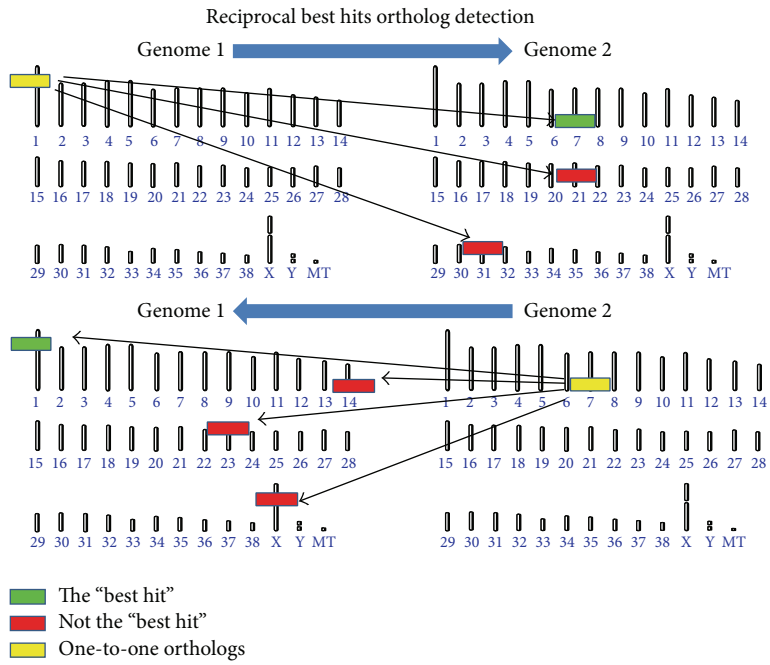


Figure 2: Identifying one-to-one orthologs between two genomes. The foundation of comparative genomics rests upon the ability to leverage annotation in a well-annotated genome in order to make inferences about genes in an undercharacterized genome. High sequence identity exists between orthologs (yellow genes, best hits) and paralogs (red genes, nonbest hits); however, paralogs are known to diverge in function much faster than orthologs. In order to achieve high confidence genotype-phenotype relationships via comparative genomics, it is essential to differentiate between orthologs and paralogs. Although complex relationships exist between orthologs, such as many-to-many and one-to-many/many-to-one, the most reliable annotation will be derived from one-to-one orthologs. The reciprocal best hit method for identifying orthologs successfully identifies one-to-one orthologs between two genomes if and only if the orthologs are the top hits to each.

mouse counterparts offers evidence for their involvement in related phenotypes within the captive population.

Unlike domesticated species, such as cats and dogs, endangered species are at risk of extinction [48], and therefore an urgency to incorporate high confidence predictions in the management is justifiable. It is worthwhile to mention that, unlike the commercial veterinary environment where, for example, proven canine genetic diagnostics may be sold to enhance the clinical management of a pet, there is relatively little financial incentive to develop commercial genetic diagnostics for endangered species as their population size is unlikely to provide an acceptable economic return on the initial investment [49].

Another critical distinction to point out is that predicting the increased susceptibility for a particular undesirable phenotype is not the same as stating a particular member of the species which will eventually have the undesirable phenotype. Rather, it is the first step in a bidirectional communication exchange between scientists and personnel managing the species (SSP managers, zoo veterinarians, and zoo animal caretakers) (Figure 5). Management plans for endangered species rely upon multiple stake holders including veterinarians, conservation organizations, and zoos [50]. In particular, the value of the phenotype predictions is immediately realized at the level of the individual member of the endangered captive species, even if the prediction turns out to be a false positive. As an illustration, consider an endangered species in which a subset of the population has been annotated as potentially having an increased risk for certain phenotypes. If this information is made available to the zoo veterinarian, these "*at-risk*" annotated animals may receive additional scrutiny when presenting with signs associated with the predicted clinical phenotype. For example, an animal that is predicted to have an increased risk for bladder cancer may be more likely to benefit from earlier detection of disease if zoo keepers are aware that an increased rate of urination may be indicative of a problem. Regardless of whether or not the animal actually has the disease, the management is informed by the genomics knowledge, which ultimately effectively triages members of a captive population in a way that maximizes their care, health, and well-being in captivity.

Just as the sharing of genomics information from the scientist to the zoo provides a more informed management of the endangered population [3, 51], sharing of the diagnosed phenotype information from the zoo veterinarian to the scientist provides feedback on the predictions and allows the scientist to refine and ultimately identify those predicted genotype-phenotype associations for which multiple clinical assessments have provided statistically significant evidence of a true genotype-phenotype relationship. Because each zoo that contains members of a captive population can contribute to the clinically relevant phenotyping, open communication among scientists, veterinarians, and zoo staff forms the basis of a genetic association study within the captive population

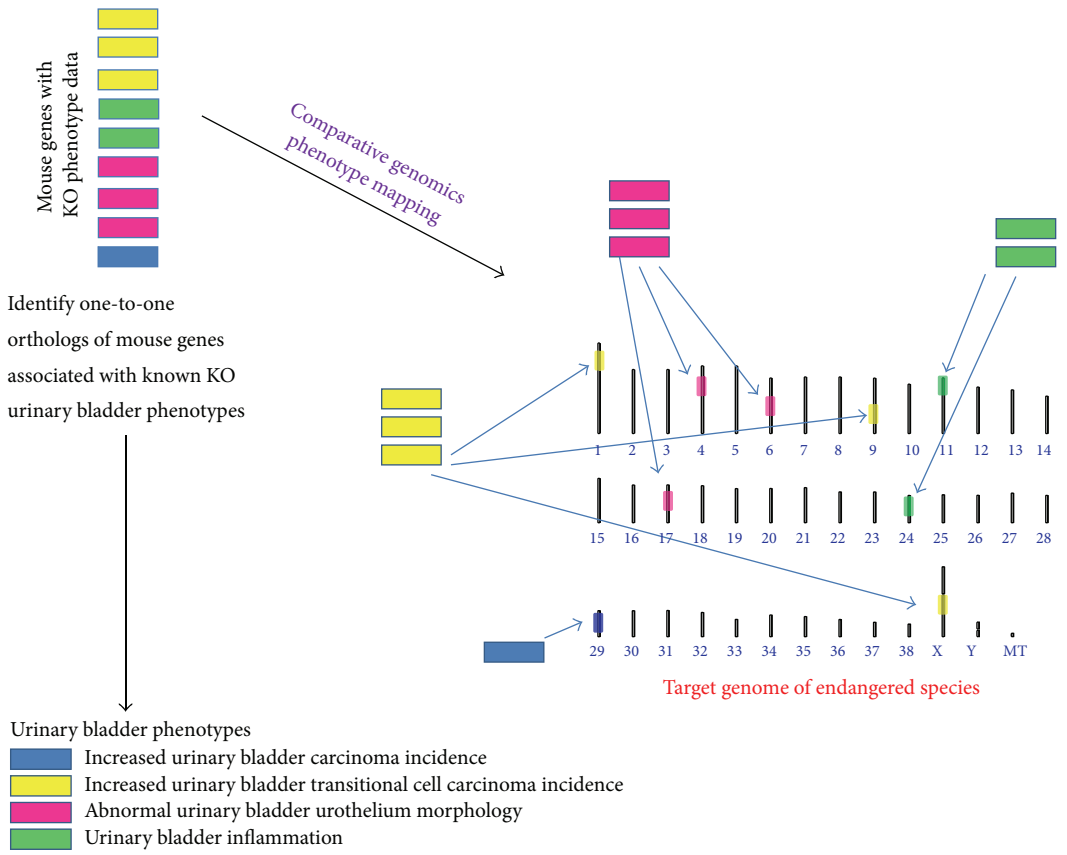


Figure 3: Comparative genomics approach to mapping phenotypes. Functional genomics information within a model organism genome, such as the mouse genome, can provide relevant and high confidence information about genes in other mammalian genomes. Nine mouse genes associated with phenotype data obtained from knockout (KO) strains of mice are represented in the top left corner of Figure 3. The resulting bladder cancer related phenotypes are indicated by the colors of the individual genes and the corresponding colors on the lower left corner of Figure 3. The mouse genes are mapped to their one-to-one orthologs in the target genome of an endangered mammalian species. The resulting locations of the orthologous genes are annotated with the same phenotype as the mouse gene. Blue: *increased urinary bladder carcinoma incidence*; yellow: *increased urinary bladder transitional carcinoma incidence*; pink: *abnormal urinary bladder urothelium morphology*; and green: *urinary bladder inflammation*. Once the phenotypes are mapped to the target genome, SNPs most likely to disrupt the orthologs in the endangered species (nonsense mutations and frameshift mutations) provide potential genotype-phenotype associations that mirror the phenotypes observed in knockout mice.

[52]. If every member of an endangered captive population is classified as either susceptible or nonsusceptible to a specific phenotype, based on bioinformatics based predictions of functional polymorphism consequences, it becomes possible to assess the increased relative risk (if any) associated with that particular genotype or allele.

In this particular "*clinical-management*" application of genotype-phenotype predictions, the information is only used to more effectively identify and medically manage the members of the captive population. Moreover, at this stage, the information does not need to be included in the breeding program. In fact, individual SSPs can decide when (if ever) to include predicted/verified genotype-phenotype relationships in the selection of breeding pairs. One plan for incorporating genotype-phenotype relationships in the breeding program might be based on the validation of the association based on multiple years of zoo veterinarians treating susceptible and nonsusceptible individuals. Threshold parameters for inclusion in a breeding program might include the calculated increased relative risk of the clinical phenotype associated with the genotype. Alternatively, the decision can consider the overall prevalence of the clinical phenotype within the captive population as well as the implications for population members that acquire the undesirable phenotype.

5. Predictive Modeling of Prevalence and Allele Frequency to Infer Relative Risk

The potential value in using genomic information to identify individuals at risk for an undesirable clinical phenotype depends upon the overall prevalence of the undesirable phenotype (disease) in the total captive population, as well as the prevalence of the susceptibility genotype (exposed) and the prevalence of the disease in the individuals with the susceptibility genotypes (disease in exposed group). As an example,

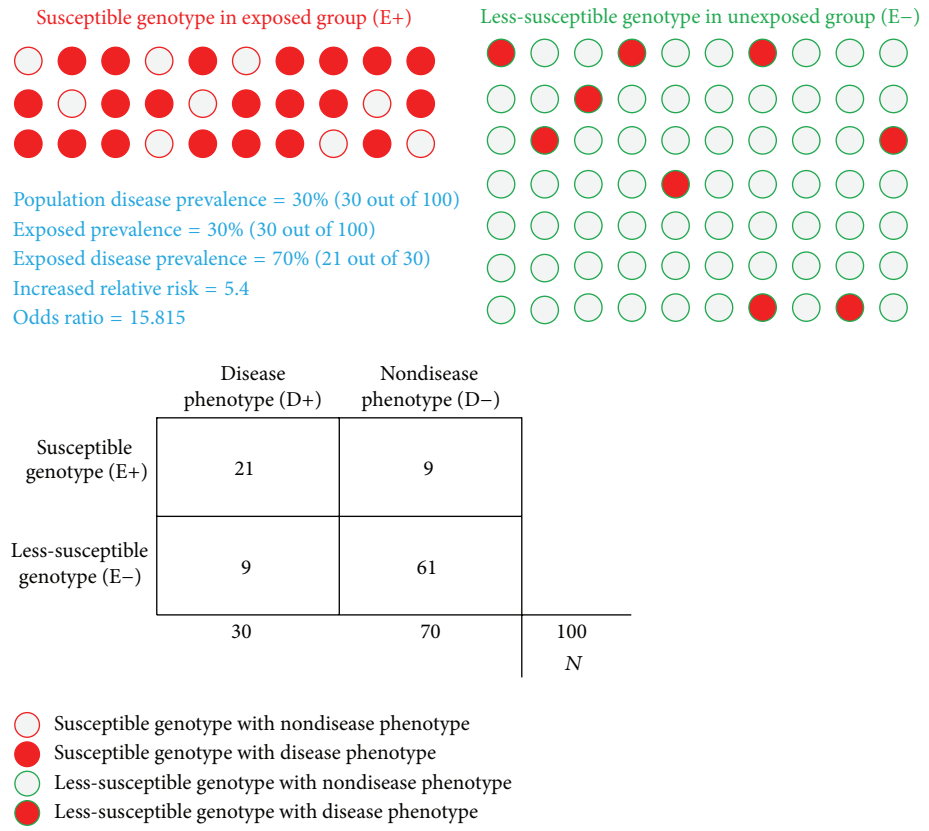


Figure 4: Predictive modeling of disease prevalence and allele frequencies. The value in using genomic information to predict individuals at risk for an undesirable clinical phenotype depends upon the prevalence of the undesirable phenotype (disease) in the total captive population, the prevalence of the susceptibility genotype (exposed), and the prevalence of the disease in the individuals with the susceptibility genotypes (disease in exposed group). The 2 × 2 matrix in the figure provides example values for a hypothetical endangered captive population of 100 individuals for which a disease phenotype has a population prevalence of 30%. The abbreviations "E+" and "E−" correspond to exposed and nonexposed, respectively. In this model, the exposed group has the allele/genotype(s) associated with the undesirable phenotype, while the nonexposed group does not have the susceptibility allele. Genotype-phenotype predictions, based on drastic SNP occurrence in genes associated with comparative genomics phenotype annotations derived from knockout mouse models, allow classification of members of the endangered species population into either a susceptible or less-susceptible class. Through bidirectional communication among zoo veterinarians, SSPs, zoo staff, and genomics scientists, genotype-phenotype predictions may be validated. Threshold values for increased relative risk in the exposed group, along with threshold levels of allele/genotype frequencies in the E+ and E− groups, will affect the success in employing such an approach for captive species management, as will the mode of inheritance (e.g., autosomal recessive versus autosomal dominant).

consider an endangered captive population of 100 individuals, for which the overall prevalence of disease is 30%, the prevalence of the susceptibility genotype is 30%, and the prevalence of the disease within the exposed group is 70% (Figure 4). This population would have an increased relative risk of 5.4 within the exposed group compared to the nonexposed group. This model assumes that some individuals without the susceptibility genotype(s) also have the undesirable phenotype, which more appropriately models polygenic and complex genetic traits. Similarly, this model assumes that some individuals with the susceptibility genotype do not have the undesirable phenotype due to incomplete penetrance. In this scenario, the 95% confidence interval for increased relative risk of the exposed group is 3.55 to 8.35 while the overall population relative risk is 2.33. Likewise the exposed group's odds ratio is 15.82 with the 95% confidence interval of 5.54 to 45.13 and the overall population odds ratio is 2.91.

A considerable number of studies investigating the relationships between allele frequency and disease in domestic animal species have been reported recently. These reports provide a framework for considering the relationship between inherited disease phenotypes and their relationship with specific alleles and/or genotypes. For example, one study investigated a specific mutation in SOD1 and the prevalence of canine degenerative myelopathy in a population of German shepherd dogs concluded that the association of the allele with the disease supported genetic testing in clinical applications [53]. The study results showed that 8 of 50 dogs exhibited homozygosity and additional 19 dogs were heterozygous.

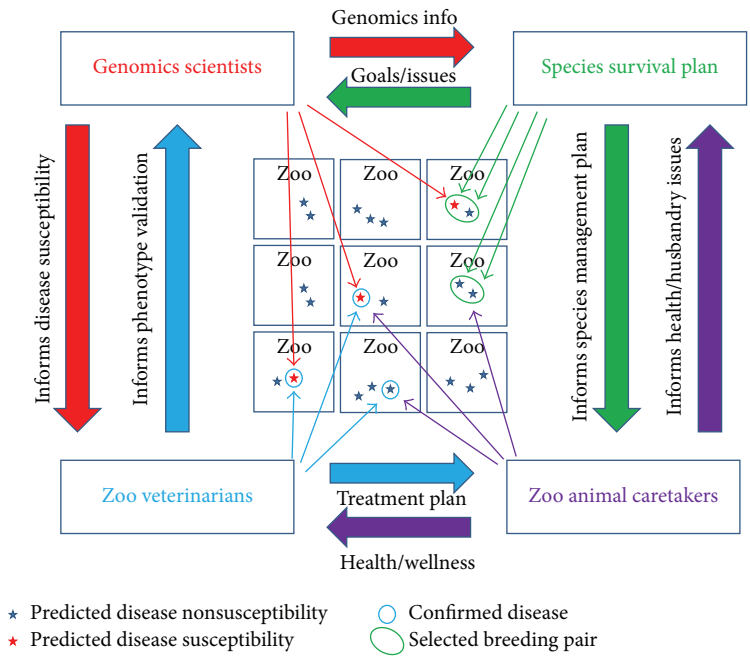


Figure 5: Bidirectional communication validates genomics predictions and enhances management. A communication network consisting of genomics scientists (red), species survival plan members (green), zoo animal caretakers (purple), and zoo veterinarians (blue) is illustrated schematically showing some of the possible communication paths and associated role of stakeholders in the network. Genomics scientists can provide genomics knowledge in which specific members of the endangered captive population are annotated as "*susceptible*" to an undesirable clinically relevant phenotype (red stars). Members of the population that are considered to be less susceptible are shown as well (dark blue stars). Red arrows originating from the genomics scientists and pointing at members of the captive population within zoos (red stars) represent the genomic knowledge applied to the captive population. Zoo veterinarians provide clinical assessment of phenotypes (blue circles around either red or dark blue stars) and subsequently validate genomics predictions. Species survival plan members select breeding pairs (green circle surrounding two stars) which can be informed by genomics information. For example, animals that are carriers for autosomal recessive undesirable phenotypes can be bred with partners that do not contain the undesirable allele, thereby allowing maximized genetic diversity while simultaneously minimizing the production of offspring with undesirable phenotypes. Zoo animal caretakers interact with the animals on a daily basis and provide and implement treatment and husbandry plans as well as serving as the eyes and ears for the endangered species. Purple arrows originating from the zoo animal caretakers and pointing to stars represent daily interactions of the zoo staff with the animals in the capacity of health, husbandry, breeding, socialization, and enrichment. The large colored arrows between the four stake holder's boxes represent examples of the types of bidirectional communication that can occur within the network to maximize the value of all information for the benefit of the endangered captive population.

Of the dogs homozygous for the undesirable allele that were greater than 8 years old, 42% exhibited a pelvic limb ataxia phenotype while none of nonataxic dogs were homozygous.

Another study investigating a retinal degenerative disease allele in 41 cat breeds determined that the undesirable allele frequency ranged from a high frequency of 33% to a low frequency of 2% in the 16 breeds in which it was detected. Clinical evaluations demonstrated a high correlation between the allele and the pathological phenotype. The authors conclude that, in breeds with the highest allele frequency, 7 to 13% of the individuals within the breed would be expected to develop the disease [54].

Finally, a 2010 study of polysaccharide storage myopathy in horses, caused by a mutation that had been identified in over 20 different horse breeds, determined that the prevalence of inherited susceptibility to the undesirable phenotype varied within these breeds from high prevalence of 62% to low prevalence of 0.5%. The management implications for this genetic information included (1) strategies for breed associations to consider screening for this specific mutation and (2) use of the undesirable allele as an "alert" for veterinarians to more closely evaluate a horse for myopathy related clinical signs such as altered gait, muscle pain, and rhabdomyolysis [55].

In these scenarios, the relationship between allele frequency and disease prevalence (or incidence) depends upon the mode of inheritance. In practice, the threshold limit of minor allele frequency in the population as well as the population disease prevalence will affect the ability to successfully employ genotype-phenotype predictions in endangered captive populations. Since the clinical assessment of undesirable phenotypes is required to validate a particular prediction, the frequency with which members of the species are clinically assessed each year will contribute to the rate at which validation can occur. Ultimately, for some phenotypes in certain species, this approach may not be viable; however,

true value in this approach lies in cases where an autosomal recessive genotype-phenotype relationship can be established in the endangered captive population. This knowledge can be applied to more effectively manage breeding by selecting breeding pairs that minimize the production of the undesirable phenotype by allowing only propagation of carriers. This "*carrier-only*" approach can maximize genetic diversity in the captive population by including affected individuals in breeding programs, especially if they are prolific breeders, while minimizing the production of affected offspring produced by these carriers of undesirable traits.

Ultimately, genomic resources coupled with per-individual genomic fingerprints can lead directly to highly improved endangered species management. For example, deleterious mutations can be identified by examining how each SNP and indel impact the gene in which the mutation is found. Highly impactful SNPs, such as those interrupting a coding region, can be mapped on a per-individual basis. If phenotypes are available in sufficient numbers of individuals, these mappings can be used to map genomic traits to the phenotype of interest using techniques such as whole genome association studies (GWAS). Where phenotype data is not sufficiently available, these mappings can be used to infer likely nondesirable configurations and used in a similar fashion on a per-individual basis. For example, if a gene has been identified and predicted to provide a phenotypically relevant function, variation mappings across individuals can be used to separate those likely healthy from those likely unhealthy. Such results can be directly applied to development of individualized health plans and to guide breeding strategy.

6. Paradigm Shift in Management Culture

The challenge in effectively applying genomics knowledge to the management of endangered captive species arises through a combination of various stake holder opinions and assumptions coupled with the limitations posed by a small highly prized population, for which individuals are easily perceived as devalued via undesirable health labels (i.e., phenotypes). Part of the issue lies in the perceived value or quality of a particular member of the species. For example, if a zoo seeks to participate in an SSP's breeding program for a particular species, acquiring an animal that may produce less desirable offspring may be problematic. Currently, an animal may be included or excluded from a breeding program based on the SSP's determination of its impact on genetic diversity of the population as determined by mean kinship calculations [56]. It also may be excluded from breeding due to the presence of an undesirable phenotype or chromosomal abnormalities [57]. If genomic-based information was to identify a greater number of individuals with a genotype linked to that phenotype, some stakeholders fear that more genetically valuable animals would be excluded from the breeding population. However, it is not required (or even preferable) to remove such "undesirable" genetic variants from the population. The power of genomics in captive endangered species may be to identify targeted breeding strategies that would minimize the impact of such variants phenotypically while still maintaining acceptable genetic diversity. In fact, removing all individuals with undesirable phenotypes from the breeding pool can result in considerable loss of genetic diversity and may even permanently remove low frequency polymorphisms of high value from the endangered captive population.

In order to maximize the value of genomics information in the management of captive endangered species, all stakeholders must appreciate the utility genomics knowledge can bring to management decisions. Successful application of genomics information in captive management will require a shift in management culture to integrate the use of the burgeoning information from the variety of resources now available to those entrusted with the care of these animals. Just as important, in order to achieve the maximum benefit that genomic information can offer, bidirectional communication must occur between caretakers managing the health of the endangered animals and scientists mining the genetic information. A shared database containing phenotypic information for each species, including physical characteristics and variations, physiologic parameters and ranges, and disease prevalence and expression, would be invaluable for functional genomics investigations. Similar databases have been previously developed for use in the breeding of transgenic mice [58]. Ultimately, conservation efforts will benefit from a prioritized commitment to value shared knowledge among all stakeholders.

The benefits of using genomics can enhance the role of zoos and SSPs by giving us knowledge about the health and attributes of a specific animal much earlier in that animal's life than our current and traditional process of waiting for and observing traits and symptoms only when they become physically apparent [59]. This can be tremendously beneficial to breeding programs as more will be known about each animal at a much earlier point in their breeding career, allowing for healthier matches and better long term genetic outcomes for the species.

Endangered species are the ideal species for the use of genomics knowledge for a number of reasons. There is tremendous public interest in and support for endangered species and the programs to preserve them. Therefore utilizing new science-based approaches, especially those that have no negative impact on the endangered individuals themselves, will enhance the public standing of the facilities and groups involved. By illustrating the value of genomics information in endangered species and the management decisions their protection requires, we can shift the paradigm for many conservation stakeholders and ultimately benefit many species fairly rapidly.

7. Conclusion

The successful conservation of wild and captive endangered species will undoubtedly evolve as genomics knowledge becomes more widely applied to management decisions. Functional genomics, as applied to conservation genetics in animal populations, entails understanding how the genome of an individual animal or the collective genomic properties of a population influences the well-being and survival of that individual or population. Genomics knowledge, such as sequence data, polymorphism data, and gene expression

data, provides an unprecedented opportunity to consider the entire genome to better understand how genetics contributes survival for a particular species. Just as comparative physiology and comparative anatomy offer a context for appreciating the mechanisms underlying variation in form and function, comparative genomics aids in elucidating conserved and divergent genetic mechanisms associated with specific phenotypes. Of particular interest are comparative genomics approaches that can efficiently and effectively identify functionally important genomic regions with implications for health and disease. Model organism genomes, from animal species, such as the mouse, dog, and cat, have already contributed to advances in the management of endangered felids and canids. As bioinformatics algorithms and pipelines become more sophisticated, the identification of genetic variants within genes implicated in specific phenotypes will facilitate identifying members of the endangered captive population that might be at risk for clinically relevant phenotypes. Such individuals can be given additional medical scrutiny to maximize the opportunity for early detection of clinically important conditions. Additionally, this genomic information can be used to better manage breeding programs, for example, to limit the production of homozygous autosomal recessive undesirable phenotypes. However, to most effectively maximize the value of genomics information in the management of captive endangered species, all stakeholders must appreciate the utility genomics knowledge can bring to management decisions. Successful application of genomics information in captive management will require a shift in management culture to integrate the use of the burgeoning information from the variety of resources now available to those entrusted with the care of these animals.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by the Institute of Museum and Library Services grant LG-54-0068-09, a generous Stromberg Award, and publication support from the College of Veterinary Medicine, Western University of Health Sciences. Kristopher J. L. Irizarry would like to specifically acknowledge the early mentorship of Catherine Plumstead who helped foster his appreciation of science, nature, and animals early during his education at Horizons on the Hudson through numerous hands on biology experiences and interactions with many animals, large and small.

References

[1] R. Ekblom and J. B. W. Wolf, "A field guide to whole-genome sequencing, assembly and annotation," *Evolutionary Applications*, vol. 7, no. 9, pp. 1026–1042, 2014.

[2] N. J. Ouborg, C. Pertoldi, V. Loeschcke, R. K. Bijlsma, and P. W. Hedrick, "Conservation genetics in transition to conservation genomics," *Trends in Genetics*, vol. 26, no. 4, pp. 177–187, 2010.

[3] C. C. Steiner, A. S. Putnam, P. E. A. Hoeck, and O. A. Ryder, "Conservation genomics of threatened animal species," *Annual Review of Animal Biosciences*, vol. 1, pp. 261–281, 2013.

[4] Z. Jiang, H. Wang, J. J. Michal et al., "Genome wide sampling sequencing for SNP genotyping: methods, challenges and future development," *International Journal of Biological Sciences*, vol. 12, no. 1, pp. 100–108, 2016.

[5] A. B. Shafer, J. B. Wolf, P. C. Alves et al., "Genomics and the challenging translation into conservation practice," *Trends in Ecology & Evolution*, vol. 30, no. 2, pp. 78–87, 2015.

[6] F. Tekaia, "Inferring orthologs: open questions and perspectives," *Genomics Insights*, vol. 9, pp. 17–28, 2016.

[7] B. J. Mcmahon, E. C. Teeling, and J. Höglund, "How and why should we implement genomics into conservation?" *Evolutionary Applications*, vol. 7, no. 9, pp. 999–1007, 2014.

[8] M. H. Kohn, W. J. Murphy, E. A. Ostrander, and R. K. Wayne, "Genomics and conservation genetics," *Trends in Ecology and Evolution*, vol. 21, no. 11, pp. 629–637, 2006.

[9] M. N. Romanov, M. Koriabine, M. Nefedov, P. J. de Jong, and O. A. Ryder, "Construction of a California condor BAC library and first-generation chicken-condor comparative physical map as an endangered species conservation genomics resource," *Genomics*, vol. 88, no. 6, pp. 711–718, 2006.

[10] M. N. Romanov, E. M. Tuttle, M. L. Houck et al., "The value of avian genomics to the conservation of wildlife," *BMC Genomics*, vol. 10, supplement 2, article S10, 2009.

[11] R. G. Felton, C. C. Steiner, B. S. Durrant, D. H. Keisler, M. R. Milnes, and C. W. Tubbs, "Identification of California condor estrogen receptors 1 and 2 and their activation by endocrine disrupting chemicals," *Endocrinology*, vol. 156, no. 12, pp. 4448–4457, 2015.

[12] B. Ujvari and K. Belov, "Major Histocompatibility Complex (MHC) markers in conservation biology," *International Journal of Molecular Sciences*, vol. 12, no. 8, pp. 5168–5186, 2011.

[13] J. B. Hans, A. Haubner, M. Arandjelovic et al., "Characterization of MHC class II B polymorphism in multiple populations of wild gorillas using non-invasive samples and next-generation sequencing," *American Journal of Primatology*, vol. 77, no. 11, pp. 1193–1206, 2015.

[14] S. A. Ellis and J. A. Hammond, "The functional significance of cattle major histocompatibility complex class I genetic diversity," *Annual Review of Animal Biosciences*, vol. 2, pp. 285–306, 2014.

[15] J. H. Robinson and A. A. Delvig, "Diversity in MHC class II antigen presentation," *Immunology*, vol. 105, no. 3, pp. 252–262, 2002.

[16] J. E. Deakin and K. Belov, "A comparative genomics approach to understanding transmissible cancer in Tasmanian devils," *Annual Review of Genomics and Human Genetics*, vol. 13, pp. 207–222, 2012.

[17] T. Callicrate, R. Dikow, J. W. Thomas, J. C. Mullikin, E. D. Jarvis, and R. C. Fleischer, "Genomic resources for the endangered Hawaiian honeycreepers," *BMC Genomics*, vol. 15, no. 1, article 1098, 2014.

[18] C. Der Sarkissian, L. Ermini, M. Schubert et al., "Evolutionary genomics and conservation of the endangered Przewalski's horse," *Current Biology*, vol. 25, no. 19, pp. 2577–2583, 2015.

[19] R. A. Griffiths and L. Pavajeau, "Captive breeding, reintroduction, and the conservation of amphibians," *Conservation Biology*, vol. 22, no. 4, pp. 852–861, 2008.

[20] G. Harding, R. A. Griffiths, and L. Pavajeau, "Developments in amphibian captive breeding and reintroduction programs," *Conservation Biology*, vol. 30, no. 2, pp. 340–349, 2016.

[21] W. Miller, S. J. Wright, Y. Zhang, S. C. Schuster, and V. M. Hayes, "Optimization methods for selecting founder individuals for captive breeding or reintroduction of endangered species," in *Proceedings of the 15th Pacific Symposium on Biocomputing (PSB '10)*, pp. 43–53, January 2010.

[22] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," *Nature Reviews Genetics*, vol. 10, no. 1, pp. 57–63, 2009.

[23] S. Bhatia, "Genetic variation as a modifier of association between therapeutic exposure and subsequent malignant neoplasms in cancer survivors," *Cancer*, vol. 121, no. 5, pp. 648–663, 2015.

[24] S. J. O'Brien, "A role for molecular genetics in biological conservation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 13, pp. 5748–5755, 1994.

[25] J. M. Heather and B. Chain, "The sequence of sequencers: the history of sequencing DNA," *Genomics*, vol. 107, no. 1, pp. 1–8, 2016.

[26] R. J. Elshire, J. C. Glaubitz, Q. Sun et al., "A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species," *PLoS ONE*, vol. 6, no. 5, Article ID e19379, 2011.

[27] W. B. Barbazuk, S. J. Emrich, H. D. Chen, L. Li, and P. S. Schnable, "SNP discovery via 454 transcriptome sequencing," *Plant Journal*, vol. 51, no. 5, pp. 910–918, 2007.

[28] G. K. Wong, B. Liu, J. Wang et al., "A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms," *Nature*, vol. 432, no. 7018, pp. 717–722, 2004.

[29] "The Human Genome Project Completion: Frequently Asked Questions," March 2016, http://www.genome.gov/11006943.

[30] R. VanBuren, D. Bryant, P. P. Edger et al., "Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*," *Nature*, vol. 527, no. 7579, pp. 508–511, 2015.

[31] M. Baker, "Functional genomics: the changes that count," *Nature*, vol. 482, no. 7384, pp. 257–262, 2012.

[32] J.-W. Choi, X. Liao, S. Park et al., "Massively parallel sequencing of Chikso (Korean brindle cattle) to discover genome-wide SNPs and InDels," *Molecules and Cells*, vol. 36, no. 3, pp. 203–211, 2013.

[33] A. Dousse, T. Junier, and E. M. Zdobnov, "CEGA-a catalog of conserved elements from genomic alignments," *Nucleic Acids Research*, vol. 44, no. 1, pp. D96–D100, 2016.

[34] B. Grumbt, S. H. Eck, T. Hinrichsen, and K. Hirv, "Diagnostic applications of next generation sequencing in immunogenetics and molecular oncology," *Transfusion Medicine and Hemotherapy*, vol. 40, no. 3, pp. 196–206, 2013.

[35] K. E. Grogan, G. J. McGinnis, M. L. Sauther, F. P. Cuozzo, and C. M. Drea, "Next-generation genotyping of hypervariable loci in many individuals of a non-model species: technical and theoretical implications," *BMC Genomics*, vol. 17, no. 1, p. 204, 2016.

[36] S. Mishra, S. K. Singh, A. K. Munjal, J. Aspi, and S. P. Goyal, "Panel of polymorphic heterologous microsatellite loci to genotype critically endangered Bengal tiger: a pilot study," *SpringerPlus*, vol. 3, article 4, 2014.

[37] W. Wang, Y. Qiao, Y. Zheng, and M. Yao, "Isolation of microsatellite loci and reliable genotyping using noninvasive samples of a critically endangered primate, *Trachypithecus leucocephalus*," *Integrative Zoology*, 2016.

[38] M. A. Nobrega and L. A. Pennacchio, "Comparative genomic analysis as a tool for biological discovery," *Journal of Physiology*, vol. 554, no. 1, pp. 31–39, 2004.

[39] M. H. de Angelis, G. Nicholson, M. Selloum et al. et al., "Analysis of mammalian gene function through broad-based phenotypic screens across a consortium of mouse clinics," *Nature Genetics*, vol. 47, no. 9, pp. 969–978, 2015.

[40] N. Ring, T. F. Meehan, A. Blake et al., "A mouse informatics platform for phenotypic and translational discovery," *Mammalian Genome*, vol. 26, no. 9-10, pp. 413–421, 2015.

[41] C. J. Bult, J. T. Eppig, J. A. Blake, J. A. Kadin, and J. E. Richardson, "Mouse genome database 2016," *Nucleic Acids Research*, vol. 44, no. 1, pp. D840–D847, 2016.

[42] S. J. O'Brien and N. Yuhki, "Comparative genome organization of the major histocompatibility complex: lessons from the Felidae," *Immunological Reviews*, vol. 167, pp. 133–144, 1999.

[43] M. A. Menotti-Raymond and S. O'Brien, "Evolutionary conservation of ten microsatellite loci in four species of felidae," *Journal of Heredity*, vol. 86, no. 4, pp. 319–322, 1995.

[44] D. C. Salim, A. A. Akimoto, C. B. Carvalho et al., "Genetic variability in maned wolf based on heterologous short-tandem repeat markers from domestic dog," *Genetics and Molecular Research*, vol. 6, no. 2, pp. 348–357, 2007.

[45] C. R. Miller, J. R. Adams, and L. P. Waits, "Pedigree-based assignment tests for reversing coyote (*Canis latrans*) introgression into the wild red wolf (*Canis rufus*) population," *Molecular Ecology*, vol. 12, no. 12, pp. 3287–3301, 2003.

[46] J. J. Johnston and L. G. Biesecker, "Databases of genomic variation and phenotypes: existing resources and future needs," *Human Molecular Genetics*, vol. 22, no. 1, Article ID ddt384, pp. R27–R31, 2013.

[47] M. A. Care, J. R. Bradford, C. J. Needham, A. J. Bulpitt, and D. R. Westhead, "Combining the interactome and deleterious SNP predictions to improve disease gene identification," *Human Mutation*, vol. 30, no. 3, pp. 485–492, 2009.

[48] P. J. Seddon, C. J. Griffiths, P. S. Soorae, and D. P. Armstrong, "Reversing defaunation: restoring species in a changing world," *Science*, vol. 345, no. 6195, pp. 406–412, 2014.

[49] P. H. Silverman, "Commerce and genetic diagnostics," *Hastings Center Report*, vol. 25, supplement 3, pp. S15–S18, 1995.

[50] P. Kelly, D. Stack, and J. Harley, "A review of the proposed reintroduction program for the far eastern leopard (Panthera pardus orientalis) and the role of conservation organizations, veterinarians, and zoos," *Topics in Companion Animal Medicine*, vol. 28, no. 4, pp. 163–166, 2013.

[51] H. B. Shaffer, M. Gidiş, E. McCartney-Melstad et al., "Conservation genetics and genomics of amphibians and reptiles," *Annual Review of Animal Biosciences*, vol. 3, pp. 113–138, 2015.

[52] R. R. Swaisgood, F. Wei, W. J. Mcshea, D. E. Wildt, A. J. Kouba, and Z. Zhang, "Can science save the giant panda (*Ailuropoda melanoleuca*)? Unifying science and policy in an adaptive management paradigm," *Integrative Zoology*, vol. 6, no. 3, pp. 290–296, 2011.

[53] A. L. Holder, J. A. Price, J. P. Adams, H. A. Volk, and B. Catchpole, "A retrospective study of the prevalence of the canine degenerative myelopathy associated superoxide dismutase 1 mutation (SOD1:c.118G > A) in a referral population of German Shepherd dogs from the UK," *Canine Genetics and Epidemiology*, vol. 1, no. 1, p. 10, 2014.

[54] M. Menotti-Raymond, V. A. David, S. Pflueger et al., "Widespread retinal degenerative disease mutation (rdAc)

discovered among a large number of popular cat breeds," *Veterinary Journal*, vol. 186, no. 1, pp. 32–38, 2010.

[55] M. E. McCue, S. M. Anderson, S. J. Valberg et al., "Estimated prevalence of the Type 1 Polysaccharide Storage Myopathy mutation in selected North American and European breeds," *Animal Genetics*, vol. 41, no. 2, pp. 145–149, 2010.

[56] A. Vinson and M. J. Raboin, "A practical approach for designing breeding groups to maximize genetic diversity in a large colony of captive rhesus macaques (*Macaca mulatta*)," *Journal of the American Association for Laboratory Animal Science*, vol. 54, no. 6, pp. 700–707, 2015.

[57] J. Rubes, P. Musilova, and G. F. Mastromonaco, "Cytogenetics of wild and captive bred non-domestic animals," *Cytogenetic and Genome Research*, vol. 120, no. 1-2, pp. 61–68, 2008.

[58] R. Hopley and A. Zimmer, "MouseBank: a database application for managing transgenic mouse breeding programs," *Biotechniques*, vol. 30, no. 1, pp. 130–132, 2001.

[59] D. Boichard and M. Brochard, "New phenotypes for new breeding goals in dairy cattle," *Animal*, vol. 6, no. 4, pp. 544–550, 2012.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 2157494, 11 pages
http://dx.doi.org/10.1155/2016/2157494

Research Article

A Quantitative Genomic Approach for Analysis of Fitness and Stress Related Traits in a *Drosophila melanogaster* Model Population

Palle Duun Rohde,[1,2,3] Kristian Krag,[1] Volker Loeschcke,[4] Johannes Overgaard,[5] Peter Sørensen,[1] and Torsten Nygaard Kristensen[6]

[1]*Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, 8830 Tjele, Denmark*
[2]*The Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), 8000 Aarhus, Denmark*
[3]*Center for Integrative Sequencing (iSEQ), Aarhus University, 8000 Aarhus, Denmark*
[4]*Section for Genetics, Ecology and Evolution, Department of Bioscience, Aarhus University, 8000 Aarhus, Denmark*
[5]*Section of Zoophysiology, Department of Bioscience, Aarhus University, 8800 Aarhus, Denmark*
[6]*Section of Biology and Environmental Science, Department of Chemistry and Bioscience, Aalborg University, 9220 Aalborg, Denmark*

Correspondence should be addressed to Palle Duun Rohde; palle.d.rohde@mbg.au.dk

Received 21 December 2015; Accepted 29 March 2016

Academic Editor: Aritz Ruiz-González

The ability of natural populations to withstand environmental stresses relies partly on their adaptive ability. In this study, we used a subset of the *Drosophila* Genetic Reference Panel, a population of inbred, genome-sequenced lines derived from a natural population of *Drosophila melanogaster*, to investigate whether this population harbors genetic variation for a set of stress resistance and life history traits. Using a genomic approach, we found substantial genetic variation for metabolic rate, heat stress resistance, expression of a major heat shock protein, and egg-to-adult viability investigated at a benign and a higher stressful temperature. This suggests that these traits will be able to evolve. In addition, we outline an approach to conduct pathway associations based on genomic linear models, which has potential to identify adaptive genes and pathways, and therefore can be a valuable tool in conservation genomics.

1. Introduction

Understanding the genetic architecture of quantitative complex traits is a central topic in modern biology, with applications ranging within evolutionary genetics, animal and plant breeding, conservation biology, and human health. Linkage analyses and candidate gene studies have been used to gather information about the genetic basis of many complex phenotypes in a range of organisms, but usually with limited power to identify the causal loci. Linkage analyses rely on the joint inheritance of a small number of markers within families with known kinship. Candidate gene studies rely on a set of preselected genes; thus, many causal genes are likely to be missed because of incomplete genetic knowledge of the trait [1–3]. With the availability of full genome sequence data, genetic polymorphisms among individuals, or within populations, can be investigated, and genome-wide association studies (GWAS) have become increasingly popular as a tool for finding causal loci [2].

The principle of GWAS is to associate the phenotypic variation with genetic polymorphism (single-nucleotide polymorphisms (SNPs) and/or other polymorphic molecular variants) with the assumption that SNPs are in linkage disequilibrium (LD) with nearby causal variants or that the SNPs themselves are causal. For an increasing number of species, the genome has been sequenced, and combined

with advances in bioinformatics, GWAS can be performed on model- as well as nonmodel organisms [4, 5]. However, studies using model organisms, including *Drosophila melanogaster*, continue to play an important role in gaining increased knowledge about the genetic architecture of complex traits. Studies on this species have contributed to elucidating the genetic architecture of complex traits, including traits associated with stress resistance [6, 7]. The *D. melanogaster* Genetic Reference Panel (DGRP) [6] has often been used in such studies. DGRP is a genetic tool which allows researchers to assay a large number of replicated individuals with the same genotype from hundreds of independent, inbred, genome-sequenced lines.

The outcome of GWAS is typically a list of the most significant SNPs. This ignores much of the remaining variance due to the conservative statistical nature of such analyses. This is unfortunate because genetic variants with small effects are likely to be missed, and even variants with large effect may not be among the top hits [8, 9]. Although valuable, the potential gain from GWAS is therefore currently not utilized sufficiently, and alternative analytical approaches are warranted to further exploit the potential use of full genome sequence data. Approaches which utilize prior biological knowledge to group SNPs (hereafter referred to as SNP-set) have been proposed to alleviate the issue of type-I error rates in analysis of large genome data sets [8–11].

Increasing evidence from studies relying on full genome approaches indicates that most traits have very complex genetic architectures and that they are influenced by often hundreds of interacting genes, each with a small effect on the phenotype and by genetic and environmental interactions [6–8, 12–15]. Accordingly, associated variants are nonrandomly distributed across the genome and are enriched within genes that interact through pathways and biological networks [12]. Such associated variants would be captured easier if based on a SNP-set based association approach where SNPs are grouped according to their physical proximity to a gene within a pathway. It can be argued that such an approach will increase the probability of finding true associations. In addition, by reducing the number of independent tests performed from the total number of SNPs (often millions) to the number of pathways (typically thousands), less restrictive statistical corrections for type-1 errors are needed. Analyzing the combined effect of many SNPs with small effect sizes might therefore increase the probability of finding causal variants [16].

In the present study we use a subset of the DGRP (21–27 lines depending on the trait assessed) to investigate five traits related to life history and environmental stress resistance in *D. melanogaster*. We use a SNP-set approach in which we aggregate SNPs based on knowledge of biological processes (here gene ontologies (GOs)) and associate the phenotypic variation with genomic variation. We investigate whether these traits harbor genetic variation for five fitness-related phenotypes: egg-to-adult viability under benign and stressful temperatures, heat stress resistance, expression of a major heat shock protein (Hsp70), and metabolic rate. The capability of a genotype to produce an egg that successfully develops into an adult fly and the ability to withstand increasing stressful temperatures are key fitness traits of major importance for the abundance and distribution of insect species [17]. Metabolic rate is known to influence functional traits such as longevity [18–20] and resistance to desiccation and starvation [21, 22]. Knowledge of the genetic variation and identification of possible pathways involved in explaining phenotypic variation in these traits is therefore of major importance for understanding species distribution and local adaptation and genomic tools are predicted to have important applications in conservation genetics in the future [23].

2. Material and Methods

2.1. Drosophila Stocks. A subset of the *Drosophila* Genetic Reference Panel (DGRP) [6] was obtained from Bloomington *Drosophila* Stock Center (NIH P40OD018537) (see Table S1 of the Supplementary Material available online at http://dx.doi.org/10.1155/2016/2157494). The flies were maintained on a standard oatmeal-sugar-yeast-agar *Drosophila* medium at 25°C and a 12-hour light/dark cycle.

2.2. Phenotypic Assays

2.2.1. Egg-to-Adult Viability. The proportion of eggs which successfully develop to adulthood was assessed at two developmental temperatures: a benign temperature (25°C) and a lightly stressful temperature (28°C). For each line and temperature 200 eggs were collected. At benign temperature, forty eggs were placed in five vials per line, and at the high temperature twenty eggs were placed in ten vials per line. Emerging flies were counted daily until all flies had emerged. To approximate Gaussian distribution data arcsine^2 was transformed. At benign and stressful temperatures 27 and 26 lines were assayed, respectively.

2.2.2. Heat Stress Resistance. Approximately 24 females from 27 lines were placed into individual glass tubes (24 × 15 mm) with lids. Tubes were randomized and placed on a rack that was submerged into a water bath heated to 37°C. Flies were constantly monitored after submersion, and heat stress resistance was measured as the time from placement in the water bath until the animal became comatose (i.e., not moving any body parts).

2.2.3. Metabolic Rate. From the rate of CO_2-production ($\dot{V}_{CO_2}$), metabolic rate was estimated using repeated stop-flow respirometry, as described in Jensen et al. [24]. Measurements were conducted in 16 parallel metabolic chambers (glass cylinder, 20 × 70 mm) over a period of 24 hours. Measurements were performed on groups of individuals with approximately 18 five-day-old females per line per metabolic chamber. On average, nine replicates per line were obtained. Each day measurements were obtained for 13 lines (and three empty controls chambers). To avoid dehydration flies had access to a solution of 4% sugar and 2% agar (0.3 mL placed on a 15 × 15 mm paper). The estimate of standard metabolic rate was obtained using the average of the three

lowest measurements of $\dot{V}_{CO_2}$ over the 24-hour period at 25 ± 1°C (see Jensen et al. [24] for details and discussion).

The stop-flow respirometry system enabled analysis of the cumulative CO_2 production for a given period as the metabolic chambers were sequentially open (3 minutes for measurement) and closed (45 minutes while CO_2 accumulates). The system was controlled by two parallel 8-channel multiplexers (RM Gas Flow Multiplexer, Sable Systems, Las Vegas, Nevada, USA). Opening allowed airflow of CO_2-striped dry air (soda lime column removes CO_2, MERCK Millipore, Darmstadt, Germany) to flush the chambers at a fixed rate of 200 mL min^{-1} controlled by a mass flow controller (MFC 2-channel v. 1.0, Sable Systems, Las Vegas, Nevada, USA) connected to a flow controller (Side-Trak®, Sierra Instruments, Monterey, California, USA). Air leaving the metabolic chambers passed through a calcium chloride column (AppliChem, Darmstadt, Germany) before entering the CO_2 analyzer (Li-6251 CO_2 Analyzer, LI-COR Environmental, Lincoln, Nebraska, USA) to remove water. The temperature inside one metabolic chamber was registered by a data logger (iButton® Data Loggers, Maxim, Sunnyvale, California, USA) and data were extracted by OneWireViewer (Maxim, Sunnyvale, California, USA). All flies were stored in a −80°C freezer after measurements and the dry weight (Sartorius Microbalance, type MC5, accuracy ± 1 μg) was obtained after drying for 24 hours at 60°C such that $\dot{V}_{CO_2}$ could be expressed as μL CO_2 produced per hour per mg dry weight.

2.2.4. Expression of Heat Shock Protein 70 Following Heat Stress. From each of 21 DGRP lines 10 five-day-old (±24 h) adult females were transferred to plastic tubes with screw cap and exposed to 35°C for 1 hour (a temperature known to induce a heat shock response; see [25]). Flies were subsequently allowed to recover at 25°C for 1 hour before being frozen at −80°C. Hsp70 expression was quantified in three replicates of approximately 10 flies for each line by ELISA, using the monoclonal antibody 7.FB which specifically binds Hsp70 in *D. melanogaster* [26], using the procedure described in Sørensen et al. [27]. Prior to analysis the data were corrected for plate effect by equalizing the mean (the same mean on the three plates).

2.3. Quantitative Genomic Analyses

2.3.1. Genomic Data. SNP data, major inversions, and *Wolbachia* status were obtained from http://dgrp2.gnets.ncsu.edu/ [6, 28]. The complete set of DGRP lines harbors 1,496,037 polymorphic markers with a minor allele frequency > 0.05. Because we used a subset of the DGRP lines monomorphic markers were observed. These were removed prior to analyses. The number of lines assayed varied among traits; thus the number of SNPs analyzed differed as well (metabolic rate: 1,179,43 SNPs, heat stress resistance: 1,231,310 SNPs, Hsp70 expression: 1,115,889 SNPs, egg-to-adult viability at benign condition; 1,231,310 SNPs, and egg-to-adult viability at light stressful condition: 1,216,721 SNPs).

2.3.2. Quantitative Genetic Parameters. Variance components and genomic effects were estimated using the REML algorithm implemented in the Regress package [29] for R [30] by fitting the linear mixed model $\mathbf{y} = \mathbf{Xb} + \mathbf{Zg} + \boldsymbol{\varepsilon}$, where $\mathbf{y}$ was a vector of phenotypic values, $\mathbf{b}$ was a vector of fixed effects (i.e., *Wolbachia* status, five major polymorphic inversions (In(2*L*)*t*, In(2*R*)*NS*, In(3*R*)*P*, In(3*R*)*K*, and In(3*R*)*Mo*), and experimental block effects), $\mathbf{g}$ was a vector of random genomic effects, and $\boldsymbol{\varepsilon}$ was a vector of the residuals. $\mathbf{X}$ and $\mathbf{Z}$ were design matrices linking fixed and genomic effects to the observations. The genomic and residual effects were defined as $\mathbf{g} \sim N(0, \mathbf{G}\sigma^2_{\mathbf{g}})$ and $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}\sigma^2_{\boldsymbol{\varepsilon}})$. The additive genomic relationship matrix, $\mathbf{G}$, was computed based on all genomic markers as $\mathbf{G} = \mathbf{WW}'/m$, where m was the total number of markers and $\mathbf{W}$ was a centered and scaled genotype matrix (i.e., mean($\mathbf{w}_i$) = 0 and var($\mathbf{w}_i$) = 1). Each column vector of $\mathbf{W}$ was $\mathbf{w}_i = (a_i - 2p_i)/\sqrt{2p_i(1-p_i)}$, where p_i was the minor allele frequency of the ith marker and a_i was the ith column vector of the allele count matrix, $\mathbf{A}$, containing the genotypes encoded as 0, 1, or 2, counting the number of minor alleles [31].

The genomic variance captured by common SNPs was computed as $h^2_{\text{SNP}} = \sigma^2_{\mathbf{g}}/(\sigma^2_{\mathbf{g}} + \sigma^2_{\boldsymbol{\varepsilon}})$, and genomic (and raw phenotypic) correlations among the traits were computed as Spearman's rank correlation. The 95% confidence interval (CI) for h^2_{SNP} was obtained using a bootstrap procedure; observations were sampled 10,000 times with replacement obtaining the same number of observations as the true data. In each round h^2_{SNP} was estimated, and the 95% CI was then obtained as the 2.5% and 97.5% quantiles of the bootstrap samples.

2.3.3. Pathway Association. We used a pathway based approach to identify sets of SNPs with association with the traits. Firstly, SNPs were annotated to genes within a 5 KB region using variant annotation from FlyBase (v.FB5.46) [32]. Secondly, based on the gene annotation, SNPs were grouped into gene ontologies (GOs) using the annotation packages GO.db [33] and org.Dm.eg.db [34] from Bioconductor [35]. Three types of GO classes were obtained: biological processes (BP), molecular function (MF), and cellular components (CC). Sets of SNPs were only included if they contained more than 10 genes, and the number of markers was >199. The numbers of markers included in the pathway association for BP, MF, and CC were 616.010, 562.938, and 549.372 SNPs, respectively.

The pathway based approach tests whether a particular SNP-set has a more extreme signal of association than a random group of SNPs. A summary statistic (T_{sum}), measuring the degree of association of one set of SNPs, was computed as the sum of the n marker effects ($\hat{s}_i$) within the given pathway; thus $T_{\text{sum}} = \sum_{i=1}^{n} \hat{s}_i$. The marker effects ($\hat{\mathbf{s}}$) were computed from the predicted genetic effect $\hat{\mathbf{g}}$ as $\hat{\mathbf{s}} = \mathbf{W}'(\mathbf{WW}')^{-1}\hat{\mathbf{g}}$.

Using a permutation approach the observed summary statistic for a SNP-set was compared to an empirical distribution of summary statistics for a random set of SNPs of the same size. As a consequence of LD, nearby SNPs will likely be correlated; this will affect the distribution of the summary

Table 1: Diagonal elements (italicized numbers) are estimated SNP heritabilities and the 95% bootstrap confidence interval in parentheses. Off-diagonal elements are genomic and raw phenotypic correlations. Below the diagonal are the Spearman rank correlations of genomic values ($\hat{\mathbf{g}}$) with associated p values and above the diagonal are the Spearman rank correlations coefficients of line means with associated p values. Numbers in bold are correlations with a $p < 0.05$.

	Metabolic rate	Heat resistance	Hsp70	Viability benign	Viability stress
Metabolic rate	*0.53 (0.38–0.60)*	0.10 (0.66)	0.06 (0.29)	0.12 (0.59)	0.14 (0.50)
Heat resistance	0.30 (0.16)	*0.41 (0.33–0.47)*	−0.19 (0.41)	0.16 (0.42)	0.14 (0.51)
Hsp70	0.19 (0.41)	0.09 (0.70)	*0.38 (0.04–0.43)*	−0.25 (0.27)	−0.16 (0.49)
Viability benign	0.24 (0.26)	−0.19 (0.34)	**−0.49 (0.02)**	*0.73 (0.60–0.77)*	**0.70 (0.00)**
Viability stress	0.21 (0.33)	−0.03 (0.87)	**−0.44 (0.05)**	**0.72 (0.00)**	*0.76 (0.70–0.79)*

statistics. To account for this correlation structure we used a procedure where we let a vector of observed marker effects be ordered according to the SNPs physical position in the genome, which then were linked to GOs. The elements in this vector were numbered $1, 2, \ldots, N$. The permutation consisted of two steps. (1) Randomly pick an element ($\hat{\mathbf{s}}_j$) from this vector. Let this jth marker effect ($\hat{\mathbf{s}}_j$) be the first element in the permuted vector and the remaining elements ordered $\hat{\mathbf{s}}_{j+1}, \hat{\mathbf{s}}_{j+2}, \ldots, \hat{\mathbf{s}}_m, \hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2, \ldots, \hat{\mathbf{s}}_{j-1}$ according to their original numbering. Thus, all elements from the original vector were shifted to a new position and starting with $\hat{\mathbf{s}}_j$ and ending with $\hat{\mathbf{s}}_{j-1}$. The mapping of GOs was kept fixed according to original mapping. (2) A summary statistic was computed for each SNP-set based on the original mapping of GOs. Hereby, the link between SNPs and pathway was broken while retaining the correlation structure among marker effects. Steps (1) and (2) were repeated 10,000 times and from this empirical distribution of summary test statistics a p value was obtained. The empirical p value corresponds to a one-tailed test of the proportion of randomly sampled summary statistics that were larger than the observed summary statistics. We assigned individual pathways as significant if $p < 0.005$.

2.3.4. Overlapping Pathways. A consequence of the small number of lines investigated was limited statistical power to detect causal variants. Therefore, we investigated whether we observed shared patterns across traits in the rankings of the pathways. For each class of pathways (i.e., BP, MF, and CC) an incidence matrix with n rows corresponding to the number of SNP-sets and m columns corresponding to the number of traits ($m = 5$) was constructed. If the summary statistic for a SNP-set was below the threshold level (here $p < 0.05$) the corresponding element in the incidence matrix was set to 1, otherwise to 0. The observed overlap was then compared to an empirical distribution of the overlap. For a total of 10,000 times the elements within each column were permutated and the overlap among columns was recorded. The probability of the overlap was estimated under the null hypothesis of independent association among traits. We determined the empirical p value of a one-tailed test as the fraction of all random permutations that was larger than or equal to the observed overlap among traits at the 5% level.

2.3.5. Partitioning of Genetic Variance within Pathways. To dissect the genetic contribution of the associated pathways, the genetic variation within pathway was decomposed to gene level. Pathway-specific genetic effects of the x_j genes constituting the pathways ($\hat{\mathbf{f}}_f$) were computed as $\hat{\mathbf{f}}_{f,x_j} = \sum_{i=1}^{m_{x_j}} \mathbf{w}_{f,x_j,i}\hat{\mathbf{s}}_i$, where $\hat{\mathbf{s}}_i$ was the genomic effect of the ith marker computed and m_{x_j} was the number of markers within the gene x_j. Thus, if a GO has the genetic effect $\hat{\mathbf{g}}_f$ and consists of x genes, then $\hat{\mathbf{g}}_f = \sum_{i=1}^{x} \hat{\mathbf{f}}_f$. A measure of the genetic variation for each feature per gene adjusted for the number of SNPs within gene ($\mathrm{Var}\, F$) was computed as $\widehat{\mathrm{Var}}F_{x_j} = \mathrm{Var}(\hat{\mathbf{f}}_{f,x_j})/m_{x_j}$.

3. Results and Discussion

In the present study we used a subset of the DGRP to investigate whether genetic variation existed for five fitness-related phenotypes, namely, the proportion of eggs that develop to adulthood at two environmental conditions, a benign and a mildly stressful temperature, resistance to acute heat stress, induction of a major heat shock protein (Hsp70), and metabolic rate estimated from CO_2 emission rate.

We found substantial phenotypic variation for all five traits (Figure 1, Table S2). In addition, genotype-by-environmental interaction (GxE) was observed for egg-to-adult viability, as the DGRP lines were not affected equally by the higher, stressful temperature (Figure 1(a)). The phenotypic variation was decomposed into a genomic ($\sigma^2_{\mathbf{g}}$) and a residual effect ($\sigma^2_{\boldsymbol{\varepsilon}}$) from which the proportion of phenotypic variation captured by common SNPs was computed, that is, h^2_{SNP}. Metabolic rate, heat stress resistance, and Hsp70 expression all showed intermediate heritability estimates, whereas the two egg-to-adult viability traits both had high estimates (Table 1). Despite the low number of DGRP lines assayed the bootstrap CI supported the magnitude of the heritability estimates, except for Hsp70 expression (Table 1). The very broad CI for Hsp70 expression was probably a consequence of the few number of lines assayed for this trait (Table S1).

Correlation of the raw phenotypic values showed a significant correlation between the two egg-to-adult viability traits (Table 1). By correlating individual genomic effects significant negative correlations were found between expression of Hsp70 and the two egg-to-adult viability traits (Table 1), and a high positive correlation between the viability traits at the two temperature conditions was found (Table 1).

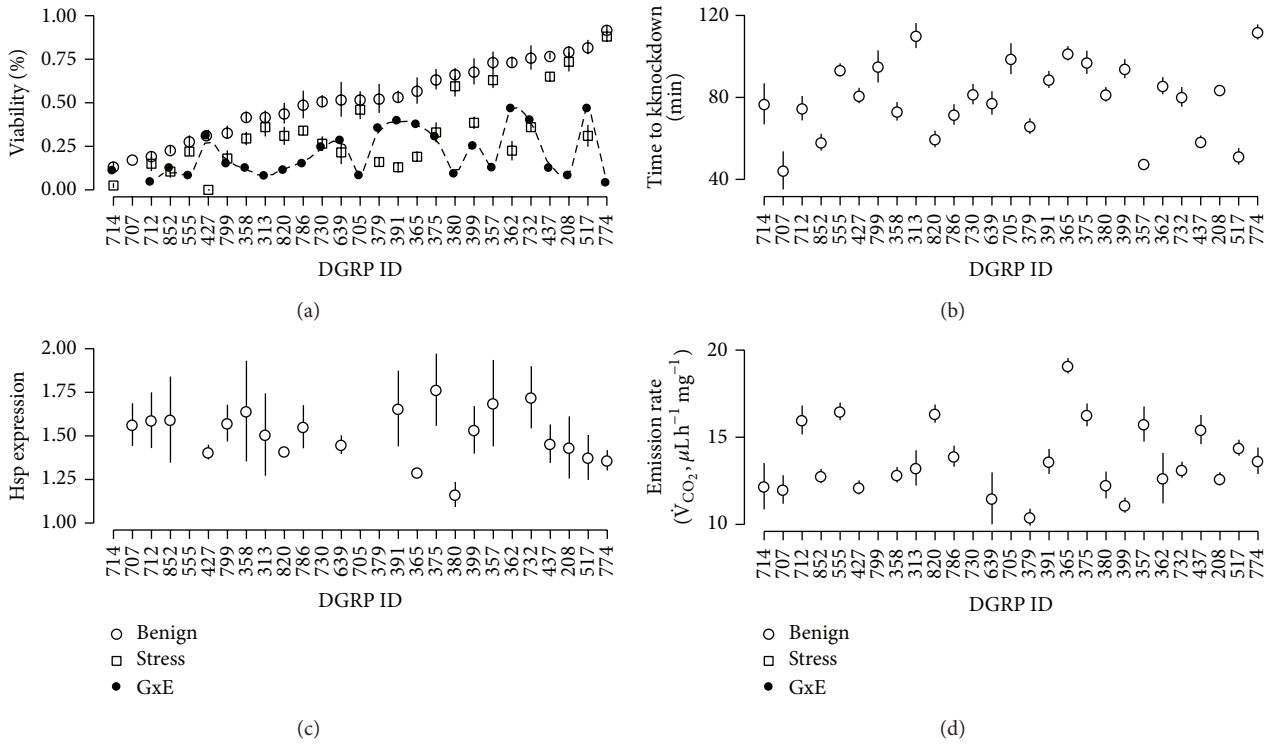


Figure 1: Distribution of DGRP phenotypes for the five assayed traits. Each panel shows the mean phenotypic value (error bars indicate standard error) for each assayed DGRP line. Lines are ordered after increasing egg-to-adult viability at benign condition. (a) Egg-to-adult viability expressed in percentage at benign condition (circles) and at lightly stressful condition (squares). Black dots indicate the difference in viability between the two environments, genotype-by-environmental interaction (GxE); (b) time to heat knockdown (min); (c) Hsp70 expression; and (d) metabolic rate measured as CO_2 emission rate.

The discrepancy between the correlations based on the raw phenotypic values and the genomic effects could be due to accounting for random and fixed effects by the mixed model. However, no strong signals were found when testing the individual fixed effects (Table S3). Infection with *Wolbachia* and the five major inversions are however known to affect trait phenotypes [28]; thus, these were kept in the model. Also, accounting for the fixed effects will take up additional degrees of freedom. However, the pathway association is not dependent on degrees of freedom; thus, this does not influence the power of the pathway test.

Based on the trait-specific individual genomic effects the trait-specific genomic marker effects were computed. Using the pathway association approach we tried, despite the rather low number of lines being assessed, to investigate whether some pathways had a more extreme signal of association than others. Pathways were divided into three categories: biological processes (BP, 689 GOs), molecular functions (MF, 239 GOs), and cellular components (CC, 161 GOs). With an arbitrary significance threshold of $p < 0.005$ the expected numbers of false-positives were three GOs in BP, one GO in MF, and below one GO in CC. For metabolic rate nine BP, two MF and four CC had a $p < 0.005$ (Table S4), which were more than expected by chance. Three MF and one BP were associated with heat stress resistance, and three of each class were associated with Hsp70 expression (Table S4). Two or fewer GO within each class were associated with the egg-to-adult viability traits (Table S4). With exception of metabolic rate, the number of GOs below the threshold value was near the expected number of false-positives. Because of the apparent limited statistical power, we investigated whether we could identify general patterns of "association" across traits. Using a less conservative threshold level we computed overlaps of GOs with a $p < 0.05$ across traits (Tables 2, 3, 4, and S5). We found statistical evidence for overlapping GOs between metabolic rate and Hsp70 expression (15 BP and four MF in common, Tables 2 and 3) and between egg-to-adult viability investigated at the two temperatures (11 BP, six MF, and four CC in common) (Tables 2, 3, and 4).

Genetic variation is necessary for populations to adapt to variable and at times stressful environments. Using a model system, we found substantial phenotypic variation for the five traits related to fitness and environmental stress resistance (Figure 1). The DGRP lines were originally inbred to an inbreeding coefficient of ~1 [6]; thus, the phenotypic variation was a consequence of genomic difference among the lines. The DGRP was established from a natural *D. melanogaster* population [6]; thus our results illustrate natural genetic variation, and therefore adaptive potential, for metabolic rate, heat stress resistance, Hsp70 expression, and egg-to-adult viability at benign and stress conditions (Table 2). Compared to, for example, morphological traits, fitness components

Table 2: Biological processes (a total of 689 GOs). Diagonal elements (italicized numbers) are the number of GOs with a $p < 0.05$. The off-diagonal elements show the number of elements shared between traits. Numbers in bold indicate a significant overlap. At a p value of 0.05 one can expect 35 false-positive SNP-sets to be assigned as significant and two SNP-sets assigned as overlapping.

	Metabolic rate	Heat resistance	Hsp70	Viability benign	Viability stress
Metabolic rate	*49*				
Heat resistance	2	*23*			
Hsp70	**15**	1	*39*		
Viability benign	4	0	3	*24*	
Viability stress	4	1	4	**11**	*30*

Table 3: Molecular function (a total of 239 GOs). Diagonal elements (italicized numbers) are the number of GOs with a $p < 0.05$. The off-diagonal elements show the number of elements shared between traits. Numbers in bold indicate a significant overlap. At a p value of 0.05 one can expect 12 false-positive SNP-sets to be assigned as significant and one SNP-set assigned as overlapping.

	Metabolic rate	Heat resistance	Hsp70	Viability benign	Viability stress
Metabolic rate	*14*				
Heat resistance	2	*12*			
Hsp70	**4**	1	*17*		
Viability benign	1	1	0	*16*	
Viability stress	2	2	0	**6**	*13*

Table 4: Cellular component (a total of 161 GOs). Diagonal elements (italicized numbers) are the number of GOs with a $p < 0.05$. The off-diagonal elements show the number of elements shared between traits. Numbers in bold indicate a significant overlap. At a p value of 0.05 one can expect eight false-positive SNP-sets to be assigned as significant and one SNP-set assigned as overlapping.

	Metabolic rate	Heat resistance	Hsp70	Viability benign	Viability stress
Metabolic rate	*11*				
Heat resistance	0	*6*			
Hsp70	2	0	*14*		
Viability benign	2	0	1	*12*	
Viability stress	1	1	2	**4**	*7*

are generally believed to have low heritability due to, for example, directional selection that removes additive genetic variation [36]. However, here we report high heritability estimates for the egg-to-adult viability traits assessed at benign and stressful temperatures. Others have reported substantial lower estimates for lifespan, fecundity, and egg-to-adult viability [37, 38]. These contradictions could result from overestimation caused by limited sample size in our study. However, the estimated heritability for Hsp70 expression was in the range of what has previously been reported [39], thereby supporting our estimates. With respect to metabolic rate our estimate was higher than other reported vales for *Drosophila* [40, 41] but was consistent with estimates obtained from bird populations [42, 43]. Lastly, the estimate for resistance to abrupt exposure to high temperatures was similar to other estimates for *Drosophila* [44] and different species of fish [45, 46]. Overall these results point to evolutionary adaptive potential for the traits investigated. Recent studies do however suggest that terrestrial ectotherms, endotherms, and plant species have limited potential to change their upper thermal limits [47, 48], which could be a consequence of hard physiological boundaries [47]. Why do our results then point to rather high evolutionary potentials in traits related to coping with high temperatures? For such comparison to be valid, the traits compared need to be similar, and it has been shown that measuring response to heat stress using a static approach (i.e., abrupt exposure to a high temperature) and that using a ramping approach (i.e., a gradual increase in temperature) are not two identical traits [49]. It has also been shown that the more ecological relevant approach (ramping) results in lower heritability estimates than the static approach [49], which we used in this study. Thus for heat resistance the results obtained in this study may not be ecologically relevant.

We found strong, significant negative genomic correlation between expression of a major heat shock protein and the two egg-to-adult viability traits and high positive genomic correlation between the two viability traits (Table 2). Egg-to-adult viability at 28°C was on average reduced by 21% compared to the viability at 25°C (Figure 1); however, the magnitude was not equal for all lines, indicative of genotype-by-environmental interaction. Studies have shown that resistance to one type of environmental stresses often has fitness costs in terms of reduced longevity or other life history traits [50–52]. Therefore, it could be hypothesized that energy spent on expression of stress response proteins, for example, Hsp70, reduces the resources available for survival. Thus,

Table 5: Genes within associated GOs that explain >20% of the genetic variation within GO.

Trait/gene ID	Gene name	Selected evidence from FlyBase [32]
Metabolic rate		
Gr28b	Gustatory receptor 28b	Feeding behavior, immune response, and thermosensory behavior
fz	Frizzled	Wnt pathway, G-protein receptor activity, and Notch signaling
bru	Brunelleschi	Meiosis cytokinesis
dl		Regulation of glucose metabolic processes, regulation of gene expression, and immune response
Fife		Regulation of neurotransmitter secretion
Heat resistance		
CG8745		Arginine catabolic processes to glutamate
CG8888		Metabolic processes
Syt1	Synaptotagmin 1	Calcium ion binding and neurotransmitter secretion
Snap25	Synaptosomal-associated protein 25 kDa	SNAP receptor activity and SNARE complex
Hsp70 expression		
fz	Frizzled	Wnt pathway, G-protein receptor activity, and Notch signaling
CG2807		Precatalytic spliceosome
Irc	Immune-regulated catalase	Response to oxidative stress
CG16941		Mitotic nuclear cell division
Pxd	Peroxidase	Response to ethanol
app	Approximated	Zinc ion binding
bin3	Bicoid-interacting protein 3	Regulation of translation
SoYb	Sister of Yb	Yb body
Viability benign		
Baldspot		Fatty-acid biosynthesis
GstS1	Glutathione S transferase S1	Glutathione metabolic process
Sply	Sphingosine-1-phosphate lyase	Sphingolipid metabolism
Viability stress		
Mkp3	Mitogen-activated protein kinase phosphatase 3	Regulation of MAPK
Timeout		DNA damage checkpoint
14-3-3ε		Determination of adult lifespan and regulation of growth

a similar correlation could be expected between metabolic rate and egg-to-adult viability, but this was not observed. However, we did observe a significant overlap in the top GOs for metabolic rate and Hsp70 expression, which does suggest genomic correlations at pathway level.

Biological interpretation from associated GOs is difficult, especially with limited statistical power to detect pathways. However, in our study, we have shown an efficient method to compute individual markers' effects and subsequently a summary statistic for a set of markers. Despite limited statistical power, which we admittedly have in this study, some biological relevant pathways were identified. For example, GO:0000149 is a group of genes related to the SNARE complex, which was associated to heat stress resistance (Table S4). Several members of the heat shock protein family are implicated in the sustenance of synaptic function, and during synaptic transmission of vesicular content, for example, neurotransmitters, heat shock proteins bind to the SNARE complex [53]. Another GO associated with heat stress was a group of genes related to oxidoreductase activity (GO:0016614, Table S4). In plants it has been shown that heat shock proteins can protect oxidoreductase complexes [23]. With respect to the GOs associated with metabolic rate several different types were identified ranging within transportation of vesicles (GO:0006891), cell division (GO:0008356), and regulation of several pathways (e.g., GO:0048786, GO:0016791, and GO:0010951). This might illustrate that metabolic rate is a complex trait influenced by many genes and that the phenotypic variation in metabolic rate is a function of variation in other biological functions within individuals.

Partitioning of genetic variation within the associated pathway to genetic variation per gene (corrected for number of genetic markers within gene) showed that most of the GOs had the same overall pattern, namely, relative few genes within each GO contributed to the overall genetic variation (Figure 2 and Table S6). Moreover, a limited number of genes explained more than 20% of the within GO genetic variation (Tables 5 and S6). Only one gene, *Frizzled*, was in common among traits, between metabolic rate and Hsp70 expression, in this reduced list of genes (Table 5). *Frizzled*

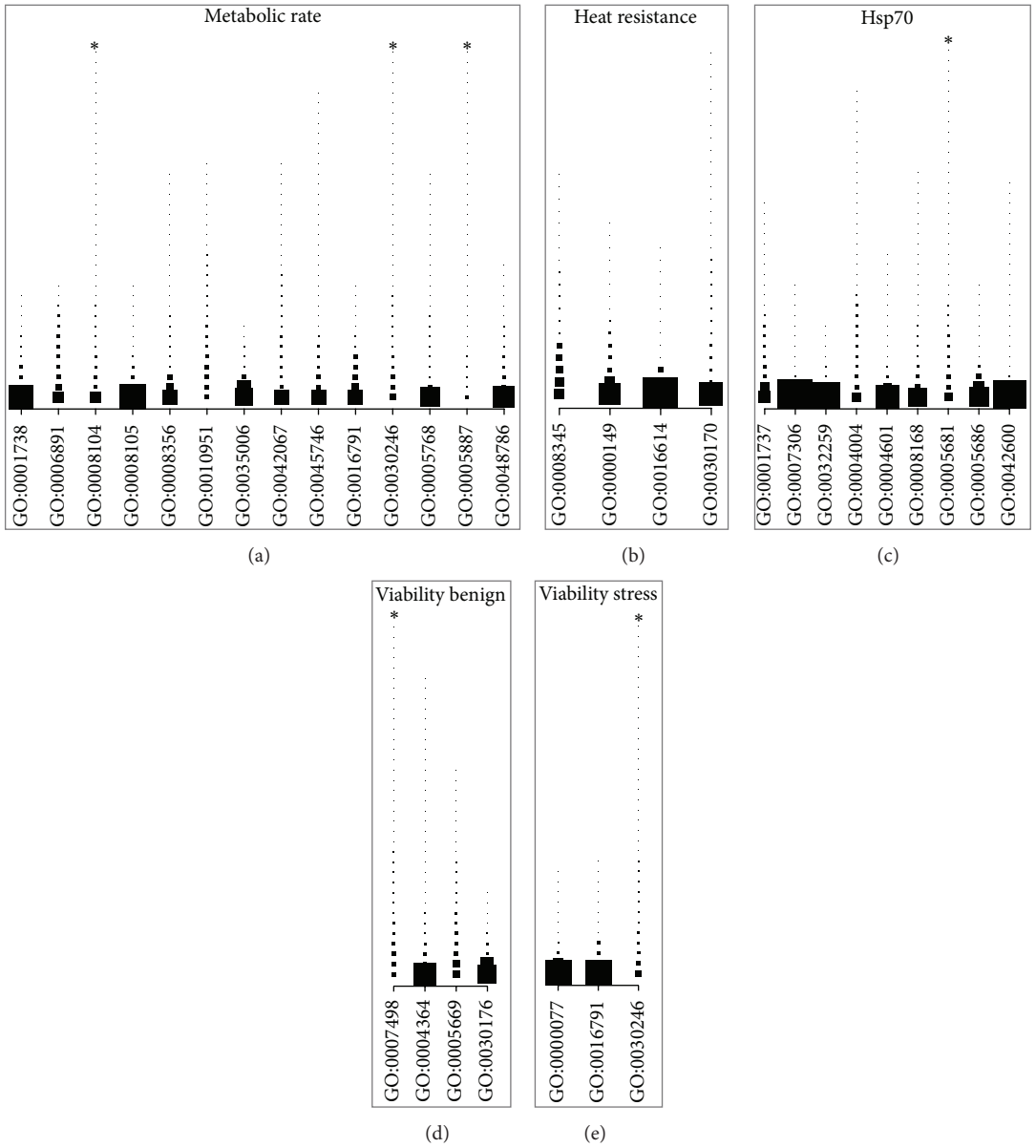


Figure 2: Partitioning of genetic variance of associated GOs to the genes constituting each GO. Proportion of variance per gene (per SNP) was standardized. Each square indicates one gene, and the size of the point indicates the relative proportion of variance explained by that gene. Asterisks (∗) indicate a truncation of the gene list. The exact values and the gene IDs can be found in Supplementary Table S6.

is associated with the Wnt signaling pathway, which promotes cell proliferation, alters key metabolic proteins, and is involved in whole-body energy homeostasis [54]. One of the genes that explained a large proportion of the genetic variation, associated with egg-to-adult viability at stressful conditions, was *Mkp3*, which is a gene in the MAPK cascade. Heat shock activates heat shock proteins which are regulated by the MAPK cascade [55]. Also, associated with heat stress resistance was *snp25*, a gene in the SNARE complex. As discussed in the previous paragraph, heat shock proteins bind to the SNARE complex. Lastly, one of the genes capturing a large fraction of the genetic variance within a GO for Hsp70 expression was *Irc*, which has been predicted to be associated with oxidative stress [32], and Hsp70 expression is known to be a biomarker for oxidative stress [56].

Despite the obvious statistical power limitations associated with the number of DGRP lines assessed in this study, we found more GOs than expected by chance, some of which seem to have important biological functions. More importantly, we found substantial genetic variation, thus evolutionary adaptive potential, for all five traits investigated supported by the relative narrow bootstrap CI for four of the traits.

With the advances in sequencing technologies genomic approaches have been suggested as promising tools for conservation genetics. Neutral markers have predominately

been applied in population and conservation genetics to describe loss of genetic variation, population structure, and so forth. However, using neutral markers to monitor the effect of environmental changes in a population is limited because the loss of variation will only decrease significantly if the population size is greatly reduced [57]. Therefore, genomic approaches, in which all polymorphic markers are used, may increase the accuracy on estimates of genetic diversity [58]. Further, with genomic approaches, it is possible to assess adaptive genes, which must harbor loci that contribute with a substantial part of the genetic variation to the traits of interest [57, 58]. Therefore, in a conservation perspective, there is a need to identify genes influencing life history and stress resistance traits. Obtaining genotypes and phenotypes of individuals from wild populations is often challenging but achievable. For organisms lacking reference sequences and proper annotations, gene regions may be predicted using traditional bioinformatic approaches. However, to achieve reliable and accurate results, the sample size must be large and therefore such approaches may not be feasible for natural populations. Thus, using genomic knowledge of key traits from model populations, such as the DGRP, to wild populations may be an alternative. Investigating and identifying genes and gene complexes associated with key traits for laboratory organisms may be used as guides for identification of variants with adaptive significance in wild populations.

4. Conclusions

In the work presented here, we used a subset of the DGRP to investigate traits related to fitness and environmental stress resistance. As the environment changes, populations need to be able to adapt to these changes to survive. Using the DGRP we found substantial genetic variation for metabolic rate, heat stress resistance, expression of Hsp70, and egg-to-adult viability at two environmental conditions. In addition, we found evidence for genotype-by-environmental interaction for viability. Using a genomic pathways association approach, we attempted to locate pathways displaying association with the traits investigated. This approach can be extended to nonmodel organisms or provided as a genomic tool for identification of adaptive genes in model organisms and thus provide a potential use for conservation genomics.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors thank Doth Andersen for technical assistance. This work was funded by the Carlsberg Foundation, the Danish Natural Research Council (Sapere Aude grants to Johannes Overgaard and Torsten Nygaard Kristensen), and the Center for Genomic Selection in Animals and Plants (GenSAP) funded by The Danish Council for Strategic Research.

References

[1] R. Patnala, J. Clements, and J. Batra, "Candidate gene association studies: a comprehensive guide to useful in silico tools," *BMC Genetics*, vol. 14, article 39, 2013.

[2] J. S. Witte, "Genome-wide association studies and beyond," *Annual Review of Public Health*, vol. 31, pp. 9–20, 2010.

[3] J. M. Kwon and A. M. Goate, "The candidate gene approach," *Alcohol Research and Health*, vol. 24, no. 3, pp. 164–168, 2000.

[4] K. Dalman, K. Himmelstrand, Å. Olson, M. Lind, M. Brandström-Durling, and J. Stenlid, "A genome-wide association study identifies genomic regions for virulence in the non-model organism *Heterobasidion annosum s.s*," *PLoS ONE*, vol. 8, no. 1, Article ID e53525, 2013.

[5] J. W. Davey, P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter, "Genome-wide genetic marker discovery and genotyping using next-generation sequencing," *Nature Reviews Genetics*, vol. 12, no. 7, pp. 499–510, 2011.

[6] T. F. C. MacKay, S. Richards, E. A. Stone et al., "The *Drosophila melanogaster* genetic reference panel," *Nature*, vol. 482, no. 7384, pp. 173–178, 2012.

[7] W. Huang, S. Richards, M. A. Carbone et al., "Epistasis dominates the genetic architecture of *Drosophila quantitative* traits," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 39, pp. 15553–15559, 2012.

[8] K. Wang, M. Li, and H. Hakonarson, "Analysing biological pathways in genome-wide association studies," *Nature Reviews Genetics*, vol. 11, no. 12, pp. 843–854, 2010.

[9] K. Wang, M. Li, and M. Bucan, "Pathway-based approaches for analysis of genomewide association studies," *American Journal of Human Genetics*, vol. 81, no. 6, pp. 1278–1283, 2007.

[10] P. Holmans, "Statistical methods for pathway analysis of genome-wide data for association with complex genetic traits," *Advances in Genetics*, vol. 72, pp. 141–179, 2010.

[11] M. Holden, S. Deng, L. Wojnowski, and B. Kulle, "GSEA-SNP: applying gene set enrichment analysis to SNP data from genome-wide association studies," *Bioinformatics*, vol. 24, no. 23, pp. 2784–2785, 2008.

[12] H. L. Allen, K. Estrada, G. Lettre et al., "Hundreds of variants clustered in genomic loci and biological pathways affect human height," *Nature*, vol. 467, pp. 832–838, 2010.

[13] P. Sarup, J. G. Sørensen, T. N. Kristensen et al., "Candidate genes detected in transcriptome studies are strongly dependent on genetic background," *PLoS ONE*, vol. 6, no. 1, Article ID e15644, 2011.

[14] T. F. C. Mackay, "Epistasis and quantitative traits: using model organisms to study gene-gene interactions," *Nature Reviews Genetics*, vol. 15, no. 1, pp. 22–33, 2014.

[15] A. J. Marian, "Molecular genetic studies of complex phenotypes," *Translational Research*, vol. 159, no. 2, pp. 64–79, 2012.

[16] M. A. Mooney, J. T. Nigg, S. K. McWeeney, and B. Wilmot, "Functional and genomic context in pathway analysis of GWAS data," *Trends in Genetics*, vol. 30, no. 9, pp. 390–400, 2014.

[17] A. A. Hoffmann and P. A. Parsons, *Evolutionary Genetics and Environmental Stress*, Oxford University Press, Oxford, UK, 1991.

[18] R. G. Melvin, W. A. Van Voorhies, and J. W. O. Ballard, "Working harder to stay alive: metabolic rate increases with age in *Drosophila simulans* but does not correlate with life span," *Journal of Insect Physiology*, vol. 53, no. 12, pp. 1300–1306, 2007.

[19] M. A. Seyahooei, J. J. M. Van Alphen, and K. Kraaijeveld, "Metabolic rate affects adult life span independently of developmental rate in parasitoid wasps," *Biological Journal of the Linnean Society*, vol. 103, no. 1, pp. 45–56, 2011.

[20] K. Niitepõld and I. Hanski, "A long life in the fast lane: positive association between peak metabolic rate and lifespan in a butterfly," *The Journal of Experimental Biology*, vol. 216, no. 8, pp. 1388–1397, 2013.

[21] M. T. Marron, T. A. Markow, K. J. Kain, and A. G. Gibbs, "Effects of starvation and desiccation on energy metabolism in desert and mesic *Drosophila*," *Journal of Insect Physiology*, vol. 49, no. 3, pp. 261–270, 2003.

[22] M. Djawdan, M. R. Rose, and T. J. Bradley, "Does selection for stress resistance lower metabolic rate?" *Ecology*, vol. 78, no. 3, pp. 828–837, 1997.

[23] C. A. Downs and S. A. Heckathorn, "The mitochondrial small heat-shock protein protects NADH: ubiquinone oxidoreductase of the electron transport chain during heat stress in plants," *FEBS Letters*, vol. 430, no. 3, pp. 246–250, 1998.

[24] P. Jensen, J. Overgaard, V. Loeschcke, M. F. Schou, H. Malte, and T. N. Kristensen, "Inbreeding effects on standard metabolic rate investigated at cold, benign and hot temperatures in *Drosophila melanogaster*," *Journal of Insect Physiology*, vol. 62, no. 1, pp. 11–20, 2014.

[25] T. N. Kristensen, J. Dahlgaard, and V. Loeschcke, "Inbreeding affects Hsp70 expression in two species of *Drosophila* even at benign temperatures," *Evolutionary Ecology Research*, vol. 4, no. 8, pp. 1209–1216, 2002.

[26] M. A. Welte, J. M. Tetrault, R. P. Dellavalle, and S. L. Lindquist, "A new method for manipulating transgenes: engineering heat tolerance in a complex, multicellular organism," *Current Biology*, vol. 3, no. 12, pp. 842–853, 1993.

[27] J. G. Sørensen, P. Michalak, J. Justesen, and V. Loeschcke, "Expression of the heat-shock protein HSP70 in *Drosophila buzzatii* lines selected for thermal resistance," *Hereditas*, vol. 131, no. 2, pp. 155–164, 1999.

[28] W. Huang, A. Massouras, Y. Inoue et al., "Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines," *Genome Research*, vol. 24, no. 7, pp. 1193–1208, 2014.

[29] D. Clifford and P. McCullagh, *The Regress Package R Package Version 1.3-14.*, 2014.

[30] R Core Team, *R: A Language and Environment for Statistical Computing*, Version 3.2.0, R Foundation for Statistical Computing, Vienna, Austria, 2015, http://www.R-project.org/.

[31] P. M. VanRaden, "Efficient methods to compute genomic predictions," *Journal of Dairy Science*, vol. 91, no. 11, pp. 4414–4423, 2008.

[32] H. Attrill, K. Falls, J. L. Goodman et al., "FlyBase: establishing a gene group resource for *Drosophila melanogaster*," *Nucleic Acids Research*, vol. 44, no. D1, pp. D786–D792, 2016.

[33] M. Carlson, *GO.db: A Set of Annotation Maps Describing the Entire Gene Ontology*, R Package Version 3.2.2, 2015.

[34] M. Carlson, org.Dm.eg.db: Genome Wide Annotation for Fly, R Package Version 3.2.3, 2015.

[35] R. C. Gentleman, V. J. Carey, D. M. Bates et al., "Bioconductor: open software development for computational biology and bioinformatics," *Genome Biology*, vol. 5, article R80, 2004.

[36] R. Frankham, J. D. Ballou, and D. A. Briscoe, *Introduction to Conservation Genetics*, Cambridge University Press, New York, NY, USA, 4th edition, 2004.

[37] M. F. Durham, M. M. Magwire, E. A. Stone, and J. Leips, "Genome-wide analysis in *Drosophila* reveals age-specific effects of SNPs on fitness traits," *Nature Communications*, vol. 5, article 4338, 2014.

[38] T. N. Kristensen, J. Overgaard, J. Lassen, A. A. Hoffmann, and C. Sgrò, "Low evolutionary potential for egg-to-adult viability in *Drosophila melanogaster* at high temperatures," *Evolution*, vol. 69, no. 3, pp. 803–814, 2015.

[39] R. A. Krebs, M. E. Feder, and J. Lee, "Heritability of expression of the 70KD heat-shock protein in *Drosophila melanogaster* and its relevance to the evolution of thermotolerance," *Evolution*, vol. 52, no. 3, pp. 841–847, 1998.

[40] P. Jumbo-Lucioni, J. F. Ayroles, M. M. Chambers et al., "Systems genetics analysis of body weight and energy metabolism traits in *Drosophila melanogaster*," *BMC Genomics*, vol. 11, article 297, 2010.

[41] A. A. Khazaeli, W. Van Voorhies, and J. W. Curtsinger, "Longevity and metabolism in *Drosophila melanogaster*: genetic correlations between life span and age-specific metabolic rate in populations artificially selected for long life," *Genetics*, vol. 169, no. 1, pp. 231–242, 2005.

[42] K. J. Mathot, K. Martin, B. Kempenaers, and W. Forstmeier, "Basal metabolic rate can evolve independently of morphological and behavioural traits," *Heredity*, vol. 111, no. 3, pp. 175–181, 2013.

[43] J.-Å. Nilsson, M. Åkesson, and J. F. Nilsson, "Heritability of resting metabolic rate in a wild population of blue tits," *Journal of Evolutionary Biology*, vol. 22, no. 9, pp. 1867–1874, 2009.

[44] O. A. Bubliy, T. N. Kristensen, V. Kellermann, and V. Loeschcke, "Humidity affects genetic architecture of heat resistance in *Drosophila melanogaster*," *Journal of Evolutionary Biology*, vol. 25, no. 6, pp. 1180–1188, 2012.

[45] C. M. Doyle, P. L. Leberg, and P. L. Klerks, "Heritability of heat tolerance in a small livebearing fish, *Heterandria formosa*," *Ecotoxicology*, vol. 20, no. 3, pp. 535–542, 2011.

[46] G. M. L. Perry, C. M. Martyniuk, M. M. Ferguson, and R. G. Danzmann, "Genetic parameters for upper thermal tolerance and growth-related traits in rainbow trout (*Oncorhynchus mykiss*)," *Aquaculture*, vol. 250, no. 1-2, pp. 120–128, 2005.

[47] M. B. Araújo, F. Ferri-Yáñez, F. Bozinovic, P. A. Marquet, F. Valladares, and S. L. Chown, "Heat freezes niche evolution," *Ecology Letters*, vol. 16, no. 9, pp. 1206–1219, 2013.

[48] A. A. Hoffmann, S. L. Chown, and S. Clusella-Trullas, "Upper thermal limits in terrestrial ectotherms: how constrained are they?" *Functional Ecology*, vol. 27, no. 4, pp. 934–949, 2013.

[49] K. A. Mitchell and A. A. Hoffmann, "Thermal ramping rate influences evolutionary potential and species differences for upper thermal limits in *Drosophila*," *Functional Ecology*, vol. 24, no. 3, pp. 694–700, 2010.

[50] O. A. Bubliy, T. N. Kristensen, V. Kellermann, and V. Loeschcke, "Plastic responses to four environmental stresses and cross-resistance in a laboratory population of *Drosophila melanogaster*," *Functional Ecology*, vol. 26, no. 1, pp. 245–253, 2012.

[51] R. A. Krebs and V. Loeschcke, "Costs and benefits of activation of the heat-shock response in *Drosophila melanogaster*," *Functional Ecology*, vol. 8, no. 6, pp. 730–737, 1994.

[52] R. Silbermann and M. Tatar, "Reproductive costs of heat shock protein in transgenic *Drosophila melanogaster*," *Evolution*, vol. 54, no. 6, pp. 2038–2045, 2000.

[53] R. A. Stetler, Y. Gan, W. Zhang et al., "Heat shock proteins: cellular and molecular mechanisms in the central nervous system," *Progress in Neurobiology*, vol. 92, no. 2, pp. 184–211, 2010.

[54] J. K. Sethi and A. Vidal-Puig, "Wnt signalling and the control of cellular metabolism," *Biochemical Journal*, vol. 427, no. 1, pp. 1–17, 2010.

[55] M. Hasanuzzaman, K. Nahar, M. M. Alam, R. Roychowdhury, and M. Fujita, "Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants," *International Journal of Molecular Sciences*, vol. 14, no. 5, pp. 9643–9684, 2013.

[56] E. El Golli-Bennour and H. Bacha, "Hsp70 expression as biomarkers of oxidative stress: mycotoxins' exploration," *Toxicology*, vol. 287, no. 1–3, pp. 1–7, 2011.

[57] A. A. Hoffmann and Y. Willi, "Detecting genetic responses to environmental change," *Nature Reviews Genetics*, vol. 9, no. 6, pp. 421–432, 2008.

[58] A. B. A. Shafer, J. B. W. Wolf, P. C. Alves et al., "Genomics and the challenging translation into conservation practice," *Trends in Ecology & Evolution*, vol. 30, no. 2, pp. 78–87, 2015.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 5304028, 7 pages
http://dx.doi.org/10.1155/2016/5304028

Review Article

The Microbiome of Animals: Implications for Conservation Biology

Simon Bahrndorff,[1] Tibebu Alemu,[2] Temesgen Alemneh,[2] and Jeppe Lund Nielsen[1]

[1]*Section of Biology and Environmental Science, Department of Chemistry and Bioscience, Aalborg University, Fredrik Bajers Vej 7H, 9220 Aalborg East, Denmark*
[2]*Department of Environmental Health Science and Technology, College of Health Science, Jimma University, Jimma, Ethiopia*

Correspondence should be addressed to Simon Bahrndorff; sba@bio.aau.dk

Received 22 January 2016; Accepted 4 April 2016

Academic Editor: Philippine Vergeer

In recent years the human microbiome has become a growing area of research and it is becoming clear that the microbiome of humans plays an important role for human health. Extensive research is now going into cataloging and annotating the functional role of the human microbiome. The ability to explore and describe the microbiome of any species has become possible due to new methods for sequencing. These techniques allow comprehensive surveys of the composition of the microbiome of nonmodel organisms of which relatively little is known. Some attention has been paid to the microbiome of insect species including important vectors of pathogens of human and veterinary importance, agricultural pests, and model species. Together these studies suggest that the microbiome of insects is highly dependent on the environment, species, and populations and affects the fitness of species. These fitness effects can have important implications for the conservation and management of species and populations. Further, these results are important for our understanding of invasion of nonnative species, responses to pathogens, and responses to chemicals and global climate change in the present and future.

1. Introduction

The microbiomes, including bacteria, fungi, and viruses, live within and upon all organisms and have become a growing area of research. With the advances of new technologies it is now possible to entangle complex microbial communities found across animal kingdoms.

Recent advances in molecular biology have provided new possibilities to investigate complex microbial communities and it has become clear that the vast majority of bacteria living in/on other animals cannot be cultured. It is now commonly accepted that at least 80% of the total bacterial species in the human gut cannot yet be cultured [1, 2].

High-throughput DNA sequencing approaches provide an attractive and cost-effective approach to investigate the composition and functions of the host microbiome. The culture-independent analysis of the host microbiome can be obtained by either metagenomic approaches or amplicon sequencing using specific marker genes. Amplicon sequencing provides a targeted version of metagenomics with a specific genetic region shared by the community members of interest. The amplified fragments derive from universal primers and are usually assumed to produce sequence read abundance that reflects the genetic diversity in the studied sample and hence sequence read abundance should reflect the genetic diversity in the studied sample. The amplified fragment typically contains phylogenetic or functional information, such as the 16S ribosomal RNA gene. 16S rRNA gene sequences are well studied and provide excellent tools for microbial community analysis [3], but other functional marker genes can also be used [4]. Subsequent taxonomy profiling of the entire microbial communities is conducted by comparisons to reference sequences or by *de novo* clustering of specific regions of sequences. Functional profiling of metagenomics is more challenging since major parts of the metagenomic data remain insufficiently characterized and

frequently samples are contaminated by host DNA or traces from the diet. Compared to both culture-dependent and more traditional molecular approaches such as sequencing of clone libraries and DGGE, amplicon sequencing approaches allow a more in depth analysis of the complete microbiome and are less restricted to the number of samples to be investigated. For further technical details see, for example, Caporaso et al. [3].

2. The Microbiome of Animals

The Human Microbiome Project (HMP) [1] was initiated in 2007 and with this it has become clear that the human microbiome is highly diverse and complex. The number of microorganisms sharing the human body is thought to outnumber human cell numbers by a factor of ten and the combined microbiome usually contains 100x more genes than its host. The microbiome also plays a major role in human health [5] and both composition and alterations in the microbiome have been found associated with diabetes, inflammatory bowel disease, obesity, asthma, rheumatoid arthritis, and susceptibility to infections [6–11].

In recent years the microbiome of a number of vertebrate nonhuman species has been sequenced including livestock [12, 13] and wildlife species such as the Tasmanian devil [14], red panda [15], giant panda [16], black howler monkey [17], and koala [18].

Insects are the most diverse and abundant groups of animals on earth [19] and have colonized many different habitats. It is therefore not surprising that insect species are also inhabited by large and diverse microbial communities playing a pivotal role for insect biology. Many insect species are inhabited by a large and diverse assembly of microorganisms, where especially the microbial communities in the intestinal tract have received much attention [20–22]. Some insect species show a much more diverse microbiome compared to other insect species. For example, the microbiomes of some synanthropic flies, such as the green bottle fly, show high diversity compared to other species such as fruit flies or mosquitoes [23–25]. The high species richness could reflect the lifestyle of synanthropic flies, for example, breeding and living by animal manure, bedding, and/or decaying organic matter rich in microorganisms.

The microbiome of other groups of invertebrates has also been established although for a limited number of species. Studies have compared the microbiome of different species of marine invertebrates with or without photosynthetic symbionts including five families of marine invertebrates [26]. Marine species of commercial interest such as oysters have also been addressed [27].

The microbes of soil invertebrates have received some attention. The gut microbes of soil animals play an indispensable role in the digestion of food and are of ecological importance in the global carbon cycle. Recently, research reported that like that of terrestrial insects some soil invertebrates such as collembolans, earthworms, and nematodes contain a rich microbiome and putative symbionts [28–30]. Further, results have shown how differences in diet among earthworm ecological groups lead to the establishment of different bacterial communities [28]. Moreover, perturbation of the soil ecosystem could impact earthworm gut wall-associated bacterial community composition and hence earthworm ecology and functioning. Even though the microbial community in invertebrates like that of collembolans and earthworms is not fully addressed, there is convincing evidence that intestinal communities can contribute to the degradation of recalcitrant biological materials such as chitin and lignocellulose [28, 29, 31].

3. Factors Affecting the Animal Microbiome and the Biological Significance

To begin with all microorganisms were seen as pathogens causing infectious diseases to the host. The host immune system of eukaryotes was built to eliminate these intruders, but at the same time tolerating its own molecules. However, we now know that the association between eukaryotic hosts and the microorganisms is far more complex. With the advances in molecular biology, such as next generation sequencing, it is now possible more specifically to address the association between a host and its microbiome. In animals the association between the host and its microbiome can take many forms and includes symbiotic and pathogenic associations [20]. Symbiotic microbiomes can be beneficial to the hosts in many ways, including dietary supplementation, host immune system, and social interactions [21, 32]. In many insects, the gut symbionts are essential for survival and development and suggest the presence of a core microbiome [33]. The symbionts need not to be completely dependent on the host and animal-microbial interactions can be flexible and facultative and the host can carry different symbionts at different times [20]. The association between the host and the microbiome is also affected by a large number of abiotic and biotic factors and can involve the immune system, nutrition, reproduction, communication, and many other systems of the host [2, 34–36].

The number of studies addressing the role of the microbiome on animal health is limited and almost entirely restricted to human studies. However, a large number of studies have addressed the role of single bacterial symbionts on animal fitness, where especially insect species have received much attention [37–39]. There is now a growing interest in understanding what factors can affect the microbiome of animals in order to understand how fitness is affected and to explain differences between ecosystems, species, and/or populations. The composition of the bacterial communities of animals including invertebrates and vertebrates seems to be shaped by multiple factors, such as the host genotype [22, 23, 40, 41], diet [17, 34, 37, 42], life stage [43], laboratory rearing [34, 43, 44], and the ecological and physiological conditions of, for example, the gut of the insect [22]. Further, recent studies have proposed that the microbiome impacts the nutritional supplementation, tolerance to environmental perturbations, and maintenance and/or development of the immune system [20].

Some invertebrates lack the complexity and diversity of associations with microorganisms. Such insect model

systems allow investigations that aim to understand the contribution of specific bacteria and the entire microbiome towards host physiological processes. For example, *Drosophila melanogaster* provide a promising model system to address some of these issues and for this species it is possible to rear axenic flies. Next generation sequencing approaches can provide an in-depth analysis of the functional roles of specific groups of bacteria and the entire microbiome on the fitness of the host. Results on *D. melanogaster* have shown how the microbiota affects developmental rate and changes metabolic rates and carbohydrate allocation under laboratory conditions [32]. Similarly functional analysis of the microbiome of ants also suggests large capacity to degrade cellulose [45] and that metabolic functions of microbes in herbivorous species play a role in fixing, recycling, or upgrading nitrogen [46]. Hypothesis has also been proposed to describe that gut microbiomes might facilitate insect herbivory and that variation in the ability to consume chemically defended plants can be partly explained by variation in the gut microbiome [47].

Recent studies have highlighted the importance of the microbiome not only in shaping the immune system but also in the context of host pathogen transmission processes (for reviews see [20, 48]). An example hereof is that the success of malaria infections is not only influenced by the mosquito innate immune responses and genetics but also affected by the composition of the gut microbiota and is in fact one of the major components affecting the outcome of mosquito infections [24]. Studies have also suggested that abiotic factors can affect the microbiome of disease vectors and thus vector competence of the host [25, 35]. Similarly the epidemics of human pathogens transmitted by insect vectors correlate with environmental factors [49, 50] suggesting that the vector competence of insect vectors is affected both indirectly and directly by environmental factors [35, 51, 52].

The recent interest in the importance of the microbiome on tolerance to environmental perturbations [38, 39] has revealed the presence of single bacterial species and mainly endosymbionts with large impact on, for example, temperature tolerance (for review see [39]). Temperature can affect the host directly or indirectly through either abundance of the symbiont or efficiency of transmission to the offspring [53–55]. At present it is unclear to what degree single strains of bacteria play a dominant role in tolerance to environmental factors or if interactions between bacteria of the microbiome are dominant. The recent advances in molecular biology and implementation of statistical analysis allow more specific hypothesis to be tested on effects of the microbiome on tolerance to, for example, environmental stress.

4. Conservation and Implications for Conservation

Changes in the microbial community have been shown to affect fitness of humans and other species as described above. However, the implications of changes in the microbiome for animal conservation have only been addressed in a limited number of studies even though the implications are many.

Several studies using next generation sequencing approaches have addressed the comparison of the microbiome of laboratory populations or individuals kept in captivity with that of wild animals [14, 15, 18, 34, 44] or of single species in habitats influenced by different degrees of human behavior [17]. Results show that species across taxa living under laboratory conditions or affected by habitat fragmentation show less diverse microbiomes compared to wild species. Thus species are jeopardized not only directly by degraded habitats with reduced resource availability but also indirectly through diminished microbiomes. It is thus essential that future studies address the microbiome and how habitat fragmentation impacts the microbiome in different species and how species with less diverse microbiomes perform under these conditions.

It is essential that we address the importance of the microbiome of other species rather than humans and the impact it has on their health status. For larger species such as primates this can be difficult and often only correlative evidence exists or can be achieved through a functional annotation of the microbiome [14, 17]. For example, in a study by Amato and coworkers [17] it was shown that beneficial fermenters, acetogens, and methanogen bacteria were more abundant in black howler monkeys inhabiting evergreen rainforest compared to individuals from fragmented habitats. The latter group also contained higher numbers of sulfate-reducing bacteria producing undesirable end products such as H_2S. This strongly suggests that habitat fragmentation will affect not only the microbiome of the host but also host fitness.

Similarly, keeping animals under captivity and maintaining breeding populations are likely to affect animal microbiomes. This is often undertaken in order to protect or increase abundance of rare species aiming at releasing species into the wild again. However, if the microbiomes of the individuals being released are affected, this is likely also to affect fitness compared to that of wild individuals and will subsequently reduce the probability of successful reintroduction into the wild. This is supported by studies on humans and mice where results have shown that obesity causes shifts in gut microbiome composition [6, 56]. Similar nutritional conditions could be expected for individuals kept in captivity. Molecular approaches allow researchers to establish entire microbiomes of animals and thus also test if, for example, it is possible to acclimate animals before being released into the wild. Optimizing environmental conditions of species in captivity could potentially ensure successful management and reintroduction.

It has been suggested that engineering microbiomes can be used to improve plant and animal health [57]. How this can be incorporated into conservation is unclear. It is standard to employ basic principles of genetics into breeding strategies for endangered species in zoos or captivity, but the microbiomes evolutionary potential has been ignored also in conservation biology.

Inbreeding has been suggested to affect the demography and persistence of natural populations and play an important role in conservation biology [58]. Recent work shows that inbreeding depression in bird and mammal populations

significantly affects birth weight, survival, reproduction, and resistance to disease, predation, and environmental stress [59]. Inbreeding depression is expected to change the proportions of homozygotes and thus also heterozygotes. Consequently recessive deleterious mutations are likely to be expressed. As fitness of animal populations is expected to be affected by genotype of the host and the microbiome and interaction between the two it is also likely that the microbiome will be affected by inbreeding depression either directly or through interaction with the genotype of the host, not only because the genepool is diminished but also because of a compromised immune system.

Microbiome analysis of wild populations has shown that the microbiome is dependent on the surrounding habitats as discussed above. This information might be used as a sensitive screening tool to establish populations affected by habitat fragmentation [17] and possibly also the effect of inbreeding. The strong signal from the diet [17, 34, 37, 42] suggests that the microbiome can also be used as a screening tool of diet preferences and to protect critical food resources or habitats for endangered species. However, it is essential that we fully understand the temporal and spatial changes in the microbiome if we are to use it as a screening tool.

The microbiome can provide protection of the host from pathogens either through stimulation of the immune system or through competitive exclusion. However, when animals are compromised or exposed to unfavorable environmental conditions the symbionts themselves can act as opportunistic pathogens [2, 27] or not provide the same degree of protection. There are examples of how environmental conditions can affect the microbiome of invertebrates. For example, studies have shown how changes in temperature have caused shifts from mutualistic to pathogen dominated communities in corals [60]. In oysters temperatures over 20°C can cause summer mortalities, but temperatures as low as 14°C will promote development of brown ring disease in clams [61, 62]. This is important in conservation biology given the fact that species and populations are or will be exposed to changes in climate under the future climate scenarios. Host species will thus be exposed to not only the direct effects of changes in, for example, temperature but also indirect effects due to change in abundance or species composition of the microbiome. These changes can again lead to direct fitness effects on the host or indirect effects through changes or modification of the immune response. The microbiome could potentially also allow organisms to respond on a short timescale and cope with, for example, changes in climate. In particular, for species with a long generation time populations might not be able to adapt to fast changes in climate. However, bacteria with a short generation time can adapt on a shorter timescale compared to the host and may provide fitness advantages that allow the host to cope with changes in climate. Future studies should more specifically test if and how the microbiome affects animals ability to respond to a changing environment. Such plastic responses can have important implications for persistence of species or populations at risk in a fluctuating environment.

Differences in microbiomes may affect invasions. For example, the interactions between native and nonnative of closely related species may be affected by the transmission of bacteria. This also appears to be associated with another emerging type of invasion, the transmission of infectious diseases of wild animals to humans [63]. Such transmission may be associated with factors including changes in human and nonhuman microbiomes. These interactions also have important implications for the conservation and management of different species within the environment. Some studies have addressed the microbiome of invasive species and also compared populations originating from the species native region with that of invasive regions [64, 65]. For the invasive snail, *Achatina fulica* results showed a highly diverse microbiome and functional analysis revealed a variety of microbial genes encoding enzymes, which is in agreement with the wide-ranging diet of this species [65]. Interestingly in another study comparing the microbiome of the soybean aphid, *Aphis glycines* from populations of native and invasive regions showed no differences [64]. Future studies should address the importance of the microbiome of invasive species to investigate if single strains of bacteria, the entire microbiome, or their interactions are major determinants for a species ability to establish in a new environment and if invasive microorganisms carried by introduced species affect native species [66].

5. Conclusions

Recent advances in molecular biology have given new possibilities to establish complex microbial communities and it has become clear that the vast majority of bacteria living in/on other animals cannot be cultured. One of the most common methods to describe complex microbiomes is the sequencing of the bacterial marker 16S ribosomal RNA (16S rRNA) genes through amplicon sequencing. Studies have shown that the microbiome plays a major role in human health, and in recent years the microbiomes of an increasing number of nonhuman species have been investigated. However, the number of studies addressing the role of the microbiome on animal health still remains limited. Some studies have discussed the role of the microbiome on nutritional supplementation, tolerance to environmental perturbations, and maintenance and development of the immune system. Thus the implications of changes in the microbiome for animal conservation are many although a limited number of studies have addressed this. We suggest that a number of factors relevant in conservation biology could affect the microbiome of animals including inbreeding, habitat fragmentation, change in climate, and effect of keeping animals in captivity. Changes in these factors are thus also likely to affect the fitness of the host both directly and indirectly. With the development of next generation sequencing and functional analysis of microbiomes it has become possible more specifically to test direct hypothesis on the importance of the microbiome in conservation biology.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the Danish Council for Independent Research (Grant no. 11-116256) (Simon Bahrndorff) and the Aalborg Zoo Conservation Foundation (AZCF).

References

[1] P. J. Turnbaugh, R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon, "The human microbiome project," *Nature*, vol. 449, no. 7164, pp. 804–810, 2007.

[2] N. Cerf-Bensussan and V. Gaboriau-Routhiau, "The immune system and the gut microbiota: friends or foes?" *Nature Reviews Immunology*, vol. 10, no. 10, pp. 735–744, 2010.

[3] J. G. Caporaso, C. L. Lauber, W. A. Walters et al., "Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms," *ISME Journal*, vol. 6, no. 8, pp. 1621–1624, 2012.

[4] M. Vital, C. R. Penton, Q. Wang et al., "A gene-targeted approach to investigate the intestinal butyrate-producing bacterial community," *Microbiome*, vol. 1, article 8, 2013.

[5] I. Cho and M. J. Blaser, "The human microbiome: at the interface of health and disease," *Nature Reviews Genetics*, vol. 13, no. 4, pp. 260–270, 2012.

[6] P. J. Turnbaugh, R. E. Ley, M. A. Mahowald, V. Magrini, E. R. Mardis, and J. I. Gordon, "An obesity-associated gut microbiome with increased capacity for energy harvest," *Nature*, vol. 444, no. 7122, pp. 1027–1031, 2006.

[7] L. Wen, R. E. Ley, P. Y. Volchkov et al., "Innate immunity and intestinal microbiota in the development of Type 1 diabetes," *Nature*, vol. 455, no. 7216, pp. 1109–1113, 2008.

[8] D. N. Frank, A. L. St Amand, R. A. Feldman, E. C. Boedeker, N. Harpaz, and N. R. Pace, "Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 34, pp. 13780–13785, 2007.

[9] J. U. Scher and S. B. Abramson, "The microbiome and rheumatoid arthritis," *Nature Reviews Rheumatology*, vol. 7, no. 10, pp. 569–578, 2011.

[10] Y. J. Huang and H. A. Boushey, "The microbiome in asthma," *Journal of Allergy and Clinical Immunology*, vol. 135, no. 1, pp. 25–30, 2015.

[11] K. Honda and D. R. Littman, "The microbiome in infectious disease and inflammation," *Annual Review of Immunology*, vol. 30, pp. 759–795, 2012.

[12] J. M. Brulc, D. A. Antonopoulos, M. E. B. Miller et al., "Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 6, pp. 1948–1953, 2009.

[13] R. Isaacson and H. B. Kim, "The intestinal microbiome of the pig," *Animal Health Research Reviews*, vol. 13, no. 1, pp. 100–109, 2012.

[14] Y. Cheng, S. Fox, D. Pemberton, C. Hogg, A. T. Papenfuss, and K. Belov, "The tasmanian devil microbiome—implications for conservation and management," *Microbiome*, vol. 3, article 76, 2015.

[15] F. Kong, J. Zhao, S. Han et al., "Characterization of the gut microbiota in the red panda (*Ailurus fulgens*)," *PLoS ONE*, vol. 9, no. 2, Article ID e87885, 8 pages, 2014.

[16] L. Zhu, Q. Wu, J. Dai, S. Zhang, and F. Wei, "Evidence of cellulose metabolism by the giant panda gut microbiome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 43, pp. 17714–17719, 2011.

[17] K. R. Amato, C. J. Yeoman, A. Kent et al., "Habitat degradation impacts black howler monkey (*Alouatta pigra*) gastrointestinal microbiomes," *The ISME Journal*, vol. 7, no. 7, pp. 1344–1353, 2013.

[18] N. Alfano, A. Courtiol, H. Vielgrader, P. Timms, A. L. Roca, and A. D. Greenwood, "Variation in koala microbiomes within and between individuals: effect of body region and captivity status," *Scientific Reports*, vol. 5, Article ID 10189, 2015.

[19] R. M. May, "How many species are there on earth?" *Science*, vol. 241, no. 4872, pp. 1441–1449, 1988.

[20] B. Weiss and S. Aksoy, "Microbiome influences on insect host vector competence," *Trends in Parasitology*, vol. 27, no. 11, pp. 514–522, 2011.

[21] A. Behar, B. Yuval, and E. Jurkevitch, "Gut bacterial communities in the Mediterranean fruit fly (*Ceratitis capitata*) and their impact on host longevity," *Journal of Insect Physiology*, vol. 54, no. 9, pp. 1377–1383, 2008.

[22] A. C.-N. Wong, J. M. Chaston, and A. E. Douglas, "The inconstant gut microbiota of *Drosophila* species revealed by 16S rRNA gene analysis," *ISME Journal*, vol. 7, no. 10, pp. 1922–1932, 2013.

[23] J. A. Chandler, J. Lang, S. Bhatnagar, J. A. Eisen, and A. Kopp, "Bacterial communities of diverse *Drosophila* species: ecological context of a host-microbe model system," *PLoS Genetics*, vol. 7, no. 9, Article ID e1002272, 2011.

[24] A. Boissière, M. T. Tchioffo, D. Bachar et al., "Midgut microbiota of the malaria mosquito vector *Anopheles gambiae* and interactions with *Plasmodium falciparum* infection," *PLoS Pathogens*, vol. 8, no. 5, Article ID e1002742, 2012.

[25] T. Wei, R. Ishida, K. Miyanaga, and Y. Tanji, "Seasonal variations in bacterial communities and antibiotic-resistant strains associated with green bottle flies (Diptera: Calliphoridae)," *Applied Microbiology and Biotechnology*, vol. 98, no. 9, pp. 4197–4208, 2014.

[26] D. G. Bourne, P. G. Dennis, S. Uthicke, R. M. Soo, G. W. Tyson, and N. Webster, "Coral reef invertebrate microbiomes correlate with the presence of photosymbionts," *ISME Journal*, vol. 7, no. 7, pp. 1452–1458, 2013.

[27] A. Lokmer and K. Mathias Wegner, "Hemolymph microbiome of Pacific oysters in response to temperature, temperature stress and infection," *ISME Journal*, vol. 9, no. 3, pp. 670–682, 2015.

[28] D. Thakuria, O. Schmidt, D. Finan, D. Egan, and F. M. Doohan, "Gut wall bacteria of earthworms: a natural selection process," *ISME Journal*, vol. 4, no. 3, pp. 357–366, 2010.

[29] T. Thimm, A. Hoffmann, H. Borkott, J. C. Munch, and C. C. Tebbe, "The gut of the soil microarthropod *Folsomia candida* (Collembola) is a frequently changeable but selective habitat and a vector for microorganisms," *Applied and Environmental Microbiology*, vol. 64, no. 7, pp. 2660–2669, 1998.

[30] N. Ladygina, T. Johansson, B. Canbäck, A. Tunlid, and K. Hedlund, "Diversity of bacteria associated with grassland soil nematodes of different feeding groups: research article," *FEMS Microbiology Ecology*, vol. 69, no. 1, pp. 53–61, 2009.

[31] M. Egert, S. Marhan, B. Wagner, S. Scheu, and M. W. Friedrich, "Molecular profiling of 16S rRNA genes reveals diet-related differences of microbial communities in soil, gut, and casts

of *Lumbricus terrestris* L. (Oligochaeta: Lumbricidae)," *FEMS Microbiology Ecology*, vol. 48, no. 2, pp. 187–197, 2004.

[32] E. V. Ridley, A. C.-N. Wong, S. Westmiller, and A. E. Douglas, "Impact of the resident microbiota on the nutritional phenotype of *Drosophila melanogaster*," *PLoS ONE*, vol. 7, no. 5, Article ID e36765, 2012.

[33] T. Hosokawa, Y. Kikuchi, N. Nikoh, M. Shimada, and T. Fukatsu, "Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria," *PLoS Biology*, vol. 4, no. 10, Article ID e337, 2006.

[34] F. Staubach, J. F. Baines, S. Künzel, E. M. Bik, and D. A. Petrov, "Host species and environmental effects on bacterial communities associated with *Drosophila* in the laboratory and in the natural environment," *PLoS ONE*, vol. 8, no. 8, Article ID e70749, 2013.

[35] R. T. Jones, R. Knight, and A. P. Martin, "Bacterial communities of disease vectors sampled across time, space, and species," *The ISME Journal*, vol. 4, no. 2, pp. 223–231, 2010.

[36] K. Lam, D. Babor, B. Duthie, E.-M. Babor, M. Moore, and G. Gries, "Proliferating bacterial symbionts on house fly eggs affect oviposition behaviour of adult flies," *Animal Behaviour*, vol. 74, no. 1, pp. 81–92, 2007.

[37] A. E. Douglas, "The microbial dimension in insect nutritional ecology," *Functional Ecology*, vol. 23, no. 1, pp. 38–47, 2009.

[38] P. Engel and N. A. Moran, "The gut microbiota of insects—diversity in structure and function," *FEMS Microbiology Reviews*, vol. 37, no. 5, pp. 699–735, 2013.

[39] H. Feldhaar, "Bacterial symbionts as mediators of ecologically important traits of insect hosts," *Ecological Entomology*, vol. 36, no. 5, pp. 533–543, 2011.

[40] J.-H. Ryu, S.-H. Kim, H.-Y. Lee et al., "Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in *Drosophila*," *Science*, vol. 319, no. 5864, pp. 777–782, 2008.

[41] V. G. Martinson, B. N. Danforth, R. L. Minckley, O. Rueppell, S. Tingek, and N. A. Moran, "A simple and distinctive microbiota associated with honey bees and bumble bees," *Molecular Ecology*, vol. 20, no. 3, pp. 619–628, 2011.

[42] B. D. Muegge, J. Kuczynski, D. Knights et al., "Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans," *Science*, vol. 332, no. 6032, pp. 970–974, 2011.

[43] Y. Wang, T. M. Gilbreath III, P. Kukutla, G. Yan, and J. Xu, "Dynamic gut microbiome across life history of the malaria mosquito *Anopheles gambiae* in Kenya," *PLoS ONE*, vol. 6, no. 9, Article ID e24767, 2011.

[44] J. L. Morrow, M. Frommer, D. C. A. Shearman, and M. Riegler, "The microbiome of field-caught and laboratory-adapted australian tephritid fruit fly species with different host plant use and specialisation," *Microbial Ecology*, vol. 70, no. 2, pp. 498–508, 2015.

[45] G. Suen, J. J. Scott, F. O. Aylward et al., "An insect herbivore microbiome with high plant Biomass-degrading capacity," *PLoS Genetics*, vol. 6, no. 9, Article ID e1001129, 2010.

[46] K. E. Anderson, J. A. Russell, C. S. Moreau et al., "Highly similar microbial communities are shared among related and trophically similar ant species," *Molecular Ecology*, vol. 21, no. 9, pp. 2282–2296, 2012.

[47] T. J. Hammer and M. D. Bowers, "Gut microbes may facilitate insect herbivory of chemically defended plants," *Oecologia*, vol. 179, no. 1, pp. 1–14, 2015.

[48] P. Azambuja, E. S. Garcia, and N. A. Ratcliffe, "Gut microbiota and parasite transmission by insect vectors," *Trends in Parasitology*, vol. 21, no. 12, pp. 568–572, 2005.

[49] S. Bahrndorff, L. Rangstrup-Christensen, S. Nordentoft, and B. Hald, "Foodborne disease prevention and broiler chickens with reduced *Campylobacte*r infection," *Emerging Infectious Diseases*, vol. 19, no. 3, pp. 425–430, 2013.

[50] B. Hald, H. M. Sommer, and H. Skovgård, "Use of fly screens to reduce *Campylobacter* spp. introduction in broiler houses," *Emerging Infectious Diseases*, vol. 13, no. 12, pp. 1951–1953, 2007.

[51] S. Bahrndorff, C. Gill, C. Lowenberger, H. Skovgard, and B. Hald, "The effects of temperature and innate immunity on transmission of *Campylobacter jejuni* (Campylobacterales: Campylobacteraceae) between life stages of *Musca domestica* (Diptera: Muscidae)," *Journal of Medical Entomology*, vol. 51, no. 3, pp. 670–677, 2014.

[52] C. C. Murdock, K. P. Paaijmans, D. Cox-Foster, A. F. Read, and M. B. Thomas, "Rethinking vector immunology: the role of environmental temperature in shaping resistance," *Nature Reviews Microbiology*, vol. 10, no. 12, pp. 869–876, 2012.

[53] A. A. Hoffmann, M. Turelli, and L. G. Harshman, "Factors affecting the distribution of cytoplasmic incompatibility in *Drosophila simulans*," *Genetics*, vol. 126, no. 4, pp. 933–948, 1990.

[54] C. B. Montllor, A. Maxmen, and A. H. Purcell, "Facultative bacterial endosymbionts benefit pea aphids *Acyrthosiphon pisum* under heat stress," *Ecological Entomology*, vol. 27, no. 2, pp. 189–195, 2002.

[55] S. S. Prado, M. Golden, P. A. Follett, M. P. Daugherty, and R. P. P. Almeida, "Demography of gut symbiotic and aposymbiotic *Nezara viridula* L. (Hemiptera: Pentatomidae)," *Environmental Entomology*, vol. 38, no. 1, pp. 103–109, 2009.

[56] R. E. Ley, F. Bäckhed, P. Turnbaugh, C. A. Lozupone, R. D. Knight, and J. I. Gordon, "Obesity alters gut microbial ecology," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 31, pp. 11070–11075, 2005.

[57] U. G. Mueller and J. L. Sachs, "Engineering microbiomes to improve plant and animal health," *Trends in Microbiology*, vol. 23, no. 10, pp. 606–617, 2015.

[58] R. Frankham, "A threshold and extinction: inbreeding," *Society for Conservation Biology*, vol. 9, pp. 792–799, 2014.

[59] L. F. Keller and D. M. Waller, "Inbreeding effects in wild populations," *Trends in Ecology and Evolution*, vol. 17, no. 5, pp. 230–241, 2002.

[60] K. B. Ritchie, "Regulation of microbial populations by coral surface mucus and mucus-associated bacteria," *Marine Ecology Progress Series*, vol. 322, pp. 1–14, 2006.

[61] C. Paillard, B. Allam, and R. Oubella, "Effect of temperature on defense parameters in Manila clam *Ruditapes philippinarum* challenged with *Vibrio tapetis*," *Diseases of Aquatic Organisms*, vol. 59, no. 3, pp. 249–262, 2004.

[62] B. T. Watermann, M. Herlyn, B. Daehne, S. Bergmann, M. Meemken, and H. Kolodzey, "Pathology and mass mortality of Pacific oysters, *Crassostrea gigas* (Thunberg), in 2005 at the East Frisian coast, Germany," *Journal of Fish Diseases*, vol. 31, no. 8, pp. 621–630, 2008.

[63] K. E. Jones, N. G. Patel, M. A. Levy et al., "Global trends in emerging infectious diseases," *Nature*, vol. 451, no. 7181, pp. 990–993, 2008.

[64] R. Bansal, M. A. R. Mian, and A. P. Michel, "Microbiome diversity of *Aphis glycines* with extensive superinfection in

native and invasive populations," *Environmental Microbiology Reports*, vol. 6, no. 1, pp. 57–69, 2014.

[65] A. M. Cardoso, J. J. V. Cavalcante, M. E. Cantão et al., "Metagenomic analysis of the microbiota from the crop of an invasive snail reveals a rich reservoir of novel genes," *PLoS ONE*, vol. 7, no. 11, Article ID e48505, 2012.

[66] P. Pyšek, V. Jarošík, P. E. Hulme et al., "A global assessment of invasive plant impacts on resident species, communities and ecosystems: the interaction of impact measures, invading species' traits and environment," *Global Change Biology*, vol. 18, no. 5, pp. 1725–1737, 2012.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 4395153, 7 pages
http://dx.doi.org/10.1155/2016/4395153

Research Article

Differential Methylation of Genomic Regions Associated with Heteroblasty Detected by M&M Algorithm in the Nonmodel Species *Eucalyptus globulus* Labill.

Rodrigo Hasbún,[1] Carolina Iturra,[1] Soraya Bravo,[2] Boris Rebolledo-Jaramillo,[3] and Luis Valledor[4]

[1]*Departamento de Silvicultura, Facultad de Ciencias Forestales, Universidad de Concepción, 4070386 Concepción, Chile*
[2]*Centro de Biotecnología Vegetal, Facultad de Ciencias Biológicas, Universidad Andrés Bello, 8370146 Santiago, Chile*
[3]*Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802, USA*
[4]*Departamento de Biología de Organismos y Sistemas, Universidad de Oviedo, 33006 Oviedo, Spain*

Correspondence should be addressed to Rodrigo Hasbún; rodrigohasbun@udec.cl

Received 24 December 2015; Revised 26 February 2016; Accepted 16 March 2016

Academic Editor: Cino Pertoldi

Epigenetic regulation plays important biological roles in plants, including timing of flowering and endosperm development. Little is known about the mechanisms controlling heterochrony (the change in the timing or rate of developmental events during ontogeny) in *Eucalyptus globulus*. DNA methylation has been proposed as a potential heterochrony regulatory mechanism in model species, but its role during the vegetative phase in *E. globulus* has not been explored. In order to investigate the molecular mechanisms governing heterochrony in *E. globulus*, we have developed a workflow aimed at generating high-resolution hypermethylome and hypomethylome maps that have been tested in two stages of vegetative growth phase: juvenile (6-month leaves) and adult (30-month leaves). We used the M&M algorithm, a computational approach that integrates MeDIP-seq and MRE-seq data, to identify differentially methylated regions (DMRs). Thousands of DMRs between juvenile and adult leaves of *E. globulus* were found. Although further investigations are required to define the loci associated with heterochrony/heteroblasty that are regulated by DNA methylation, these results suggest that locus-specific methylation could be major regulators of vegetative phase change. This information can support future conservation programs, for example, selecting the best methylomes for a determinate environment in a restoration project.

1. Introduction

Tree species are usually able to tolerate a wide range of environmental conditions. In fact, they are able to tolerate a wide range of environmental conditions and, in many cases, extreme seasonal changes [1]. Some of these organisms are able to manifest different phenotypes depending on the environment in which they grow. This phenomenon, called phenotypic plasticity, has been defined as a change in the phenotype expressed by a single genotype in different environments [2, 3]. Phenotypic plasticity evolves to maximize fitness when the environment is variable and increases with latitude [4–6].

Leaf heteroblasty is a significant and abrupt change in form and function that occurs over the maturation process (phase change from seedlings to reproductive individuals) of certain plants [7]. Characteristics commonly affected include leaf form, size, and arrangement. The earlier and later stages of leaf development are named juvenile and adult, respectively. In contrast to phenotypic plasticity, heteroblastic development does not depend on environmental cues. However, the timing or rate of heteroblastic changes, which can be referred to as a type of heterochrony—change of relative timing of events throughout development—can be modified by the environment [8]. Heterochrony has been implicated in plant evolution, because it can impact the

physiology, phytochemistry, or resistance to pests and disease of certain plants [9, 10]; however it is a plastic response largely underexplored.

Phenotypic plasticity and heterochrony may interact to produce a pattern of variation in the leaf phenotype, even in organisms with little or no genetic diversity. Most endangered species have lower genetic diversity than related nonendangered species [11], and phenotypic plasticity and heterochrony together can increase the possibilities of adaption/persistence. Therefore, knowledge of the molecular mechanisms regulating both processes could open new alternatives to assist conservation programs.

Epigenetic regulation, in particular DNA methylation, plays an important biological role in plants, including timing of flowering and endosperm development [12]. Transgenerational inheritance of DNA methylation can mediate phenotypic plasticity via novel epialleles and phenotypes within populations/species [13]. In a review by Pascual et al. [3] it was shown that the coordination of genetic and epigenetic mechanisms mediated phenotypic variation in different plants. For example, in populations of *Arabidopsis thaliana* with experimental alteration of DNA methylation, the overall patterns of variability among the genotypes indicated that epigenetic changes could affect not only the short-term environmental responses of plants, but also the evolutionary potential of important traits and their plasticity [14]. Similarly, a recent study in invertebrates proposed that the absence of germline DNA methylation in genes involved in the response to fluctuating conditions facilitates phenotypic variation, which could contribute to increased adaptive potential [15]. In conifers it has been reported that environment influences a differential DNA methylation during embryogenesis, inducing differential priming of the embryos that causes differential capabilities to adapt to environment [16].

There is little evidence linking epigenetic regulation and heterochrony. Only a few studies focusing on epigenetic changes during leaf differentiation and development have been developed in *Arabidopsis* [17–19], rice [20], or pine [21] but none of these species are strongly heteroblastic. Environmental cues are perceived as input signals for the microRNA156/SQUAMOSA promoter-binding protein-like (SPL or SBP box) module and act as a quantitative developmental clock of phase transitions in plants [22, 23]. The evidence shows that sugars promote vegetative phase change through their effect on miR156 [24], but other endogenous factors could play additional important roles. The same authors proposed that heritable epigenetic modification of the miR156 precursor and/or additional chromatin structure alterations could regulate heteroblasty [24].

Eucalyptus globulus, a tree of Australian origin but cultivated worldwide, is strongly heteroblastic with clear differences between its juvenile and adult leaves [25]. Jordan et al. [26] found that genetic association of the timing of vegetative phase change with growth rate ranged from positive to negative at different sites. Early phase change may be favored in warm, wet environments to reduce damage produced by leaf fungi, but it may also be favored on exposed dry sites to increase form or plant structure by which it is protected from desiccation [27, 28]. Genome-wide DNA methylation maps of many model organisms have been reported, but in nonmodel organisms like *Eucalyptus* spp. the methylation patterns remain poorly studied. These types of maps can be applied to a wide range of biological problems, using the analysis of methylation differences between ecotypes or individuals within species [29–31].

The aim of this study is to set up the required methodology and assess the epigenetic changes (hypermethylation and hypomethylation) related to heteroblasty in *E. globulus*. Specifically, we want to detect differentially methylated regions (DMRs) between juvenile and adult leaves of *E. globulus*. DMRs are stretches of genomic DNA that have different DNA methylation patterns. Their natural variation could guide the conservation management of the species or the selection of individuals with potentially adequate methylomes—set of modifications of nucleic acid methylation in the genome of an organism—for a discrete environmental condition. We expect that the results and experience from this work could be used for the discovery of key regulators of heterochrony in future studies, which could be used to assist conservation programs of threatened species.

2. Materials and Methods

To investigate the molecular mechanisms governing heterochrony in *Eucalyptus globulus,* we applied the M&M algorithm [32] to identify DMRs related to heteroblasty of vegetative growth. We used the previously sequenced X46 clone (Mininco SA JGI Project ID: 401875). Ten ramets of the X46 clone were produced by cuttings, and a genetic trial was established in the commune of Renaico, province of Malleco, region of La Araucanía, Chile (latitude −37.67, longitude −72.59). Juvenile leaf material from five plants was harvested after 6 months at nodes 8 to 10, mixed, and stored at −70°C. Plants were grown until the vegetative phase change was evident (after more than 2 years of growth, average 45 nodes), and adult leaf material was collected. Genomic DNA was extracted using the DNeasy Plant Mini Kit (QIAGEN Inc.).

The methylation profiles were determined by DNA sequencing of enriched genomic libraries: (i) hypermethylome (the methylated part of the genome) using immunoprecipitation of methylated DNA (MeDIP-seq) and (ii) hypomethylome (the nonmethylated part of the genome) based on restriction enzymes sensitive to methylation (MRE-seq).

2.1. MeDIP-seq and MRE-seq Library Generation and Sequencing. MeDIP-seq is a large-scale purification technique used to enrich libraries for methylated DNA sequences. It consists of isolating methylated DNA fragments via an antibody raised against 5-methylcytosine (5mC). MRE-seq utilizes a combination of methyl sensitive enzymes to enrich libraries for unmethylated DNA sequences.

Libraries were generated as previously described in Li et al. [32], with minor modifications. For MeDIP-seq, 3500 ng of isolated DNA was sonicated using 26 pulses of 10 s ON/20 s OFF (Sonic Dismembrator model 100, Fisher Scientific) to

a fragment size of 100–500 bp, end processed, and ligated to paired-end adapters using NEXTflex PCR free DNA Sequencing Kit (Bioo Scientific). After size selection of 166–566 bp using Agencourt AMPure XP (Beckman Coulter), DNA was heat denatured and then immunoprecipitated using Methylated DNA IP Kit (Zymo Research), using a mouse monoclonal anti-5-methylcytosine antibody according to manufacturer's instructions. DNA was then purified with Agencourt AMPure XP (Beckman Coulter) and eluted in 25 mL resuspension buffer (10 mM Tris-HCl, pH 8.5). DNA was amplified by 12 cycles of PCR with the standard Illumina index primers. For MRE-seq, three digestion reactions (HpaII, AciI, and Hin6I; Fermentas) were performed in parallel, each with 600 ng of DNA. Ten units of enzyme (except AciI, which uses five units) were initially incubated with DNA for 3 h, and then additional five units of enzyme were added to the digestion for a total of 6 h of digestion time. Digested DNA from the different reactions was combined and purified using ChiP DNA Clean & Concentrator™ (Zymo Research). The purified DNA was end processed and ligated to single-end adapters using NEXTflex PCR free DNA Sequencing Kit (Bioo Scientific). After size selection (166–566 bp) with Agencourt AMPure XP (Beckman Coulter), the DNA was amplified by PCR for six cycles.

MeDIP and MRE libraries were sent to the DNA Sequencing Facility of the Biotechnology Center at the University of Wisconsin. Samples were sequenced on an Illumina HiSeq machine, yielding a total of 204 million MeDIP-seq reads and 236 million MRE-seq reads. The reads were mapped to the latest *Eucalyptus grandis* genome assembly (v2.0) [33], using BWA-MEM Li [34] with default settings.

2.2. Use of M&M Algorithm to Detect DMRs. We used an algorithm named M&M [32] that integrates data from both MeDIP-seq and MRE-seq to detect DMRs. M&M is available as an R package called "methylMnM." Briefly, M&M integrates MeDIP-seq and MRE-seq by dynamically scaling, normalizing, and combining the datasets and provides exact p value and q-value for different sample comparison. The coverage of MeDIP and MRE sequencing data and genomic CpG information were calculated for each 2000 bp genomic bin. Before applying the M&M method, we generated two input files: (1) CpG sites of each window and (2) MRE-CpG sites of each window. To generate file (1) we used the script fasta2bed.py (Computational Genomics Analysis Tools CGAT 0.2.3), and to calculate (2) we used a script created specifically for this project (discussed at https://www.biostars.org/p/86480/). DMRs between developmental stages were identified using the M&M algorithm with default parameters in the R environment (version v.2.12.1). Briefly, the coverage of MeDIP and MRE sequencing data and genomic CpG information were calculated for each 2000 bp genomic bin. Scaffolds were excluded from the analysis. DMRs with a q-value of 1E − 7 were selected for analysis. Several statistical functions are implemented in the methylMnM package: "MnM.test()" to calculate the probability that the methylation levels of the two samples within each bin were different, "MnM.$q$value()" to estimate q-values based on all the p values, and "MnM.select-DMR()" to select significant DMRs based on a cutoff of q-value $<$ 1E − 4. The output files contained genomic locations of statistically significant DMRs and their MeDIP-seq and MRE-seq values (in RPKM), as well as p values and q-values. The absolute values of genomic regions are negative $\log_{10}$-transformed q-values. If the value is negative, it represents hypermethylation in the vegetative juvenile sample and hypomethylation in the vegetative adult; if the value is positive, it represents hypomethylation in the vegetative juvenile sample and hypermethylation in the vegetative adult. Mapping results and detected DMRs were visually inspected with the GenomeView software [35] using default parameters.

3. Results and Discussion

3.1. Sequencing Results. For each developmental stage, we constructed one sequencing library using two complementary technologies: MeDIP-seq and MRE-seq. These libraries were sequenced to generate 446 million reads in total (Table 1), of which 387 million were mapped to the *E. grandis* genome covering at least 50%. We note that MeDIP-seq data has less mapping efficiency. Li et al. [32] explained this because relatively more MeDIP-seq reads are derived from repetitive regions of the genome, which are often heavily methylated. Some reads from repetitive regions cannot be mapped uniquely. In line with expectations MRE-seq scores were inversely correlated with MeDIP-seq scores.

The two sequencing signals covered nonoverlapping regions as expected, but some adjacent genomics regions showed overlapping signals (Figure 1).

3.2. Identification of DMRs. The main aim of this work was the determination of the local DNA methylation changes between two developmental phases (i.e., vegetative juvenile and vegetative adult). We only considered CpG methylation and applied the M&M algorithm to our data to identify DMRs. This allowed us to find a total of 1090 putative DMRs (q-value $<$ 1E − 4) between the two developmental stages. More than 70% of DMRs showed increasing DNA methylation levels from vegetative growth phases juvenile to adult. These preliminary results suggest that locus-specific methylation patterns could be an important feature of vegetative heteroblasty control. This increment in the number of methylated regions agrees with the reduction in the number of genes and proteins that is observed in mature leaves (when they reach their full physiological competence) compared to leaves in proliferative stage [36, 37] which have been also related to an increase of global DNA methylation in conifers [21, 38] and *Arabidopsis* [39].

In future work, candidate DMRs related to genes involved in plastic responses will be validated using more genotypes of *E. globulus* with heterochronic responses in the expression of heteroblasty. Using quantitative techniques for quantifying methylation (e.g., bisulfite/sequencing of specific genomic locus) we hope to find DMRs that could explain or predict heterochronic responses. Knowledge of genomic loci that regulate heterochrony in *E. globulus* will improve our understanding of molecular mechanism of this process and can support future conservation programs, for example,

Table 1: Summary of mapping statistic from MeDIP-seq and MRE-seq libraries of juvenile and adult leaves.

	Total fragments (Tf)	Mapped fragments	% of Tf	Uniquely mapped fragments	% of Tf	Nonredundant uniquely mapped fragments	Genome coverage (%)
MeDIP-seq-juvenile	$1.37E+8$	$1.17E+8$	85.4	$6.74E+7$	49.1	$6.79E+6$	52.3
MeDIP-seq-adult	$0.70E+8$	$0.57E+8$	82.0	$2.48E+7$	35.4	$6.72E+6$	38.3
MRE-seq-juvenile	$1.14E+8$	$1.02E+8$	89.5	$7.91E+7$	69.4	$9.6E+6$	57.0
MRE-seq-adult	$1.25E+8$	$1.11E+8$	88.8	$8.83E+7$	70.6	$6.34E+6$	49.2
Total	$4.46E+8$	$3.87E+8$	86.7	$2.60E+8$	58.3	$2.95E+7$	

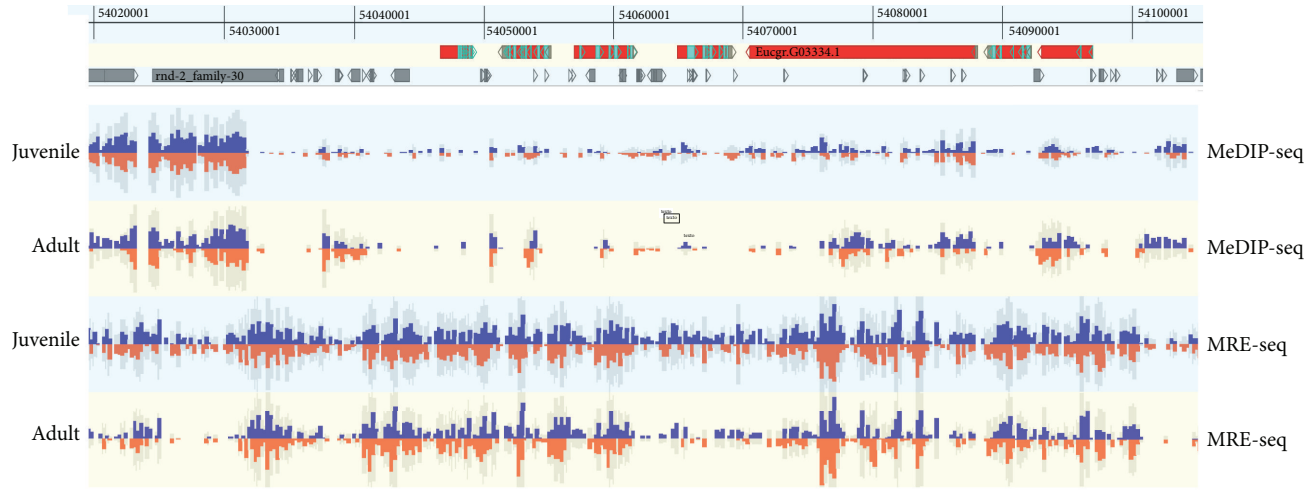


Figure 1: Epigenome view (GenomeView software) of a 138.7 kb region of chromosome 07 (54,020,001–54,100,001) of *Eucalyptus grandis* reference genome. MeDIP-seq libraries and MRE-seq libraries covered largely nonoverlapping and short overlapping regions.

selecting the best methylomes for a specific environment in a restoration project. This complete workflow could be easily applied to other nonmodel species with the only requirement of a reference genome. Given that the cost of DNA sequencing is falling [40] and bioinformatics facilities for assembly of massive sequences are increasing quickly [41], generating a reference genome for almost any species is becoming more affordable. The latter also is valid for threatened species.

The introduction of a workflow allowing the study of the epigenetic regulation at genome level will provide new insight that goes further than the mere description of the genes that are regulated during plant development or environmental adaption. For instance, having the possibility of tracking epigenetic changes in the form of metastable epialleles—alleles that are variably expressed in genetically identical individuals due to epigenetic modifications established during early development and are thought to be particularly vulnerable to environmental influences—will allow us to add a new source of variation, which is particularly important for threatened species, which usually exhibit a very low genetic diversity [42, 43]. Although these studies can be done routinely in model species [44, 45], it is particularly difficult to adapt wet and *in silico* methodologies for nonmodel like trees, since tissues are rich in polyphenols and other contaminant molecules requiring optimizations (e.g., [46–48]) and at the same time bioinformatics pipelines (gene prediction, characterization, and annotation algorithms) still require improvements [49]. In this study we describe a workflow aiming at achieving results comparable to those obtained in *Arabidopsis* in terms of quality of libraries and quantity of reads. This procedure will allow future studies employing a higher number of samples and experimental situations that permit addressing important issues related to conservation and management of threatened species such as restoration and translocation. Furthermore, the integration of this regulatory layer together with other omic levels has been proved to be useful for explaining adaptive divergences (see Meijón et al. [50] for an example of the power of high-throughput approaches to explore natural variation in tree species).

4. Conclusion

These preliminary results suggest that locus-specific methylation process could be an important feature of vegetative heteroblasty control in *E. globulus*. The workflow set up in this project opens a promising future for discovering DNA methylation patterns among different tissue types, cell types, and individuals that will help us to explain phenotypic plasticity and adaption capabilities through the basis of a divergent epigenetic regulation. High-throughput epigenomic technology and analytical tools used in this study could be applied to population-based studies of nonmodel plants but principal challenges are experimental design, data analysis, and interpretation of results. The implementation of epigenetic fingerprinting as a support tool in restoration and conservation projects of threatened species requires the discovery of loci involved in adaptive variation.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

The authors thank MININCO Forest Company for providing plant material and land for the field trial. María Francisca Beltrán, Paulina Rivas, Macarena Arellano, Giannina Espinoza, Joseline Valdes, and Daniela Fernandez are gratefully acknowledged for providing technical assistance. The authors thank the DNA Sequencing Facility of the Biotechnology Center at the University of Wisconsin for providing sequencing and support services. Funding for this

project was provided by CONICYT-Chile (Project FONDECYT INICIACION no. 11110505).

References

[1] T. W. Crowther, H. B. Glick, K. R. Covey et al., "Mapping tree density at a global scale," *Nature*, vol. 525, no. 7568, pp. 201–205, 2015.

[2] L. Gratani, "Plant phenotypic plasticity in response to environmental factors," *Advances in Botany*, vol. 2014, Article ID 208747, 17 pages, 2014.

[3] J. Pascual, M. J. Cañal, B. Correia et al., "Can epigenetics help forest plants to adapt to climate change?" in *Epigenetics in Plants of Agronomic Importance: Fundamentals and Applications*, R. Alvarez-Venegas, C. De la Peña, and J. A. Casas-Mollano, Eds., pp. 125–146, Springer, Cham, Switzerland, 2014.

[4] A. A. Agrawal, "Phenotypic plasticity in the interactions and evolution of species," *Science*, vol. 294, no. 5541, pp. 321–326, 2001.

[5] S. Bahrndorff, V. Loeschcke, C. Pertoldi, C. Beier, and M. Holmstrup, "The rapid cold hardening response of Collembola is influenced by thermal variability of the habitat," *Functional Ecology*, vol. 23, no. 2, pp. 340–347, 2009.

[6] J. Overgaard, T. N. Kristensen, K. A. Mitchell, and A. A. Hoffmann, "Thermal tolerance in widespread and tropical *Drosophila* species: does phenotypic plasticity increase with latitude?" *The American Naturalist*, vol. 178, no. 1, pp. S80–S96, 2011.

[7] G. Zotz, K. Wilhelm, and A. Becker, "Heteroblasty—a review," *The Botanical Review*, vol. 77, no. 2, pp. 109–151, 2011.

[8] S. J. Gould, *Ontogeny and Phylogeny*, Harvard University Press, Cambridge, Mass, USA, 1st edition, 1977.

[9] P. Li and M. O. Johnston, "Heterochrony in plant evolutionary studies through the twentieth century," *Botanical Review*, vol. 66, no. 1, pp. 57–88, 2000.

[10] C. J. Hudson, J. S. Freeman, R. C. Jones et al., "Genetic control of heterochrony in *Eucalyptus globulus*," *G3: Genes, Genomes, Genetics*, vol. 4, no. 7, pp. 1235–1245, 2014.

[11] R. Frankham, J. D. Ballou, and D. A. Briscoe, *Introduction to Conservation Genetics*, Cambridge University Press, Cambridge, UK, 1st edition, 2002.

[12] E. J. Finnegan, W. J. Peacock, and E. S. Dennis, "DNA methylation, a key regulator of plant development and other processes," *Current Opinion in Genetics and Development*, vol. 10, no. 2, pp. 217–223, 2000.

[13] Y.-Y. Zhang, M. Fischer, V. Colot, and O. Bossdorf, "Epigenetic variation creates potential for evolution of plant phenotypic plasticity," *New Phytologist*, vol. 197, no. 1, pp. 314–322, 2013.

[14] O. Bossdorf, D. Arcuri, C. L. Richards, and M. Pigliucci, "Experimental alteration of DNA methylation affects the phenotypic plasticity of ecologically relevant traits in *Arabidopsis thaliana*," *Evolutionary Ecology*, vol. 24, no. 3, pp. 541–553, 2010.

[15] S. B. Roberts and M. R. Gavery, "Is there a relationship between DNA methylation and phenotypic plasticity in invertebrates?" *Frontiers in Physiology*, vol. 2, pp. 116–120, 2012.

[16] I. A. Yakovlev, Y. Lee, B. Rotter et al., "Temperature-dependent differential transcriptomes during formation of an epigenetic memory in *Norway spruce* embryogenesis," *Tree Genetics and Genomes*, vol. 10, no. 2, pp. 355–366, 2014.

[17] M. Lafos, P. Kroll, M. L. Hohenstatt, F. L. Thorpe, O. Clarenz, and D. Schubert, "Dynamic regulation of H3K27 trimethylation during arabidopsis differentiation," *PLoS Genetics*, vol. 7, no. 4, Article ID e1002040, 2011.

[18] M. Lodha, C. F. Marco, and M. C. P. Timmermans, "The ASYMMETRIC LEAVES complex maintains repression of KNOX homeobox genes via direct recruitment of Polycomb-repressive complex2," *Genes and Development*, vol. 27, no. 6, pp. 596–601, 2013.

[19] J. A. Brusslan, A. M. Rus Alvarez-Canterbury, N. U. Nair, J. C. Rice, M. J. Hitchler, and M. Pellegrini, "Genome-wide evaluation of histone methylation changes associated with leaf senescence in Arabidopsis," *PLoS ONE*, vol. 7, no. 3, Article ID e33151, 2012.

[20] P. Sui, J. Jin, S. Ye et al., "H3K36 methylation is critical for brassinosteroid-regulated plant growth and development in rice," *Plant Journal*, vol. 70, no. 2, pp. 340–347, 2012.

[21] L. Valledor, J. Pascual, M. Meijón, M. Escandón, and M. J. Cañal, "Conserved epigenetic mechanisms could play a key role in regulation of photosynthesis and development-related genes during needle development of *Pinus radiata*," *PLoS ONE*, vol. 10, no. 5, Article ID e0126405, 2015.

[22] H. Wang and H. Wang, "The miR156/SPL module, a regulatory hub and versatile toolbox, gears up crops for enhanced agronomic traits," *Molecular Plant*, vol. 8, no. 5, pp. 677–688, 2015.

[23] A. Pulido and P. Laufs, "Co-ordination of developmental processes by small RNAs during leaf development," *Journal of Experimental Botany*, vol. 61, no. 5, pp. 1277–1291, 2010.

[24] L. Yang, M. Xu, Y. Koo, J. He, and R. Scott Poethig, "Sugar promotes vegetative phase change in *Arabidopsis thaliana* by repressing the expression of *MIR156A* and *MIR156C*," *eLife*, vol. 2013, no. 2, Article ID e00260, 2013.

[25] C. J. Hudson, J. S. Freeman, R. C. Jones et al., "Genetic control of heterochrony in *Eucalyptus globulus*," *G3: Genes|Genomes|Genetic*, vol. 4, no. 7, pp. 1235–1245, 2014.

[26] G. J. Jordan, B. M. Potts, P. Chalmers, and R. J. E. Wiltshire, "Quantitative genetic evidence that the timing of vegetative phase change in *Eucalyptus globulus* ssp. globulus is an adaptive trait," *Australian Journal of Botany*, vol. 48, no. 5, pp. 561–567, 2000.

[27] G. W. Dutkowski and B. M. Potts, "Geographic patterns of genetic variation in *Eucalyptus globulus* ssp. globulus and a revised racial classification," *Australian Journal of Botany*, vol. 47, no. 2, pp. 237–263, 1999.

[28] G. J. Jordan, B. M. Potts, and R. J. E. Wiltshire, "Strong, independent, quantitative genetic control of the timing of vegetative phase change and first flowering in *Eucalyptus globulus* ssp. globulus," *Heredity*, vol. 83, no. 2, pp. 179–187, 1999.

[29] N. A. Eckardt, "Deep sequencing maps the maize epigenomic landscape," *Plant Cell*, vol. 21, no. 4, pp. 1024–1026, 2009.

[30] F. Roudier, F. K. Teixeira, and V. Colot, "Chromatin indexing in *Arabidopsis*: an epigenomic tale of tails and more," *Trends in Genetics*, vol. 25, no. 11, pp. 511–517, 2009.

[31] X. Li, J. Zhu, F. Hu et al., "Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression," *BMC Genomics*, vol. 13, article 300, 2012.

[32] D. Li, B. Zhang, X. Xing, and T. Wang, "Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation," *Methods C*, vol. 72, pp. 29–40, 2015.

[33] J. Bartholomé, E. Mandrou, A. Mabiala et al., "High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly," *New Phytologist*, vol. 206, no. 4, pp. 1283–1296, 2015.

[34] H. Li and R. Durbin, "Fast and accurate short read alignment with Burrows-Wheeler transform," *Bioinformatics*, vol. 25, no. 14, pp. 1754–1760, 2009.

[35] T. Abeel, T. Van Parys, Y. Saeys, J. Galagan, and Y. Van de Peer, "GenomeView: a next-generation genome browser," *Nucleic Acids Research*, vol. 40, no. 2, p. e12, 2012.

[36] L. Valledor, J. V. Jorrín, J. L. Rodríguez et al., "Combined proteomic and transcriptomic analysis identifies differentially expressed pathways associated to *Pinus radiata* needle maturation," *Journal of Proteome Research*, vol. 9, no. 8, pp. 3954–3979, 2010.

[37] M. Andriankaja, S. Dhondt, S. DeBodt et al., "Exit from proliferation during leaf development in *Arabidopsis thaliana*: a not-so-gradual process," *Developmental Cell*, vol. 22, no. 1, pp. 64–78, 2012.

[38] L. Valledor, M. Meijón, R. Hasbún, M. Jesús Cañal, and R. Rodríguez, "Variations in DNA methylation, acetylated histone H4, and methylated histone H3 during *Pinus radiata* needle maturation in relation to the loss of in vitro organogenic capability," *Journal of Plant Physiology*, vol. 167, no. 5, pp. 351–357, 2010.

[39] E. J. Finnegan, W. J. Peacock, and E. S. Dennis, "Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 16, pp. 8449–8454, 1996.

[40] K. A. Wetterstrand, "DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)," 2016, http://www.genome.gov/sequencingcosts/.

[41] R. Ekblom and J. B. W. Wolf, "A field guide to whole-genome sequencing, assembly and annotation," *Evolutionary Applications*, vol. 7, no. 9, pp. 1026–1042, 2014.

[42] N. J. Ouborg, P. Vergeer, and C. Mix, "The rough edges of the conservation genetics paradigm for plants," *Journal of Ecology*, vol. 94, no. 6, pp. 1233–1248, 2006.

[43] R. Frankham, "Challenges and opportunities of genetic approaches to biological conservation," *Biological Conservation*, vol. 143, no. 9, pp. 1919–1927, 2010.

[44] M. J. Ziller, H. Gu, F. Müller et al., "Charting a dynamic DNA methylation landscape of the human genome," *Nature*, vol. 500, no. 7463, pp. 477–481, 2013.

[45] S. R. Eichten, R. Briskine, J. Song et al., "Epigenetic and genetic influences on DNA methylation variation in maize populations," *Plant Cell*, vol. 25, no. 8, pp. 2783–2797, 2013.

[46] R. Hasbún, L. Valledor, J. L. Rodríguez et al., "HPCE quantification of 5-methyl-2′deoxycytidine in genomic DNA: methodological optimization for chestnut and other woody species," *Plant Physiology and Biochemistry*, vol. 46, no. 8-9, pp. 815–822, 2008.

[47] L. Valledor, R. Hasbún, R. Rodríguez, and M. J. Cañal, "RNA-Free DNA extraction protocol from pinus tissues for molecular biology or HPCE/HPLC analyses," *Journal of Plant Biochemistry and Biotechnology*, vol. 18, no. 2, pp. 229–232, 2009.

[48] R. Hasbún, C. Iturra, P. Moraga et al., "An efficient and reproducible protocol for production of AFLP markers in tree genomes using fluorescent capillary detection," *Tree Genetics and Genomes*, vol. 8, no. 4, pp. 925–931, 2012.

[49] P. Radivojac, W. T. Clark, T. R. Oron et al., "A large-scale evaluation of computational protein function prediction," *Nature Methods*, vol. 10, no. 3, pp. 221–227, 2013.

[50] M. Meijón, I. Feito, M. Oravec et al., "Exploring natural variation of *Pinus pinaster* Aiton using metabolomics: is it possible to identify the region of origin of a pine from its metabolites?" *Molecular Ecology*, vol. 25, no. 4, pp. 959–976, 2016.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 5613862, 9 pages
http://dx.doi.org/10.1155/2016/5613862

Research Article

The Use of Genomics in Conservation Management of the Endangered Visayan Warty Pig (*Sus cebifrons*)

Rascha J. M. Nuijten,[1,2] Mirte Bosse,[1] Richard P. M. A. Crooijmans,[1] Ole Madsen,[1] Willem Schaftenaar,[3] Oliver A. Ryder,[4] Martien A. M. Groenen,[1] and Hendrik-Jan Megens[1]

[1]*Animal Breeding and Genomics Centre (ABGC), Wageningen University, 6708 PB Wageningen, Netherlands*
[2]*Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), 6708 PB Wageningen, Netherlands*
[3]*Veterinary Services, Rotterdam Zoo, 3041 JG Rotterdam, Netherlands*
[4]*San Diego Zoo Institute for Conservation Research, Escondido, CA 92027, USA*

Correspondence should be addressed to Rascha J. M. Nuijten; rascha.nuijten@gmail.com

Received 24 December 2015; Accepted 7 February 2016

Academic Editor: Cino Pertoldi

The list of threatened and endangered species is growing rapidly, due to various anthropogenic causes. Many endangered species are present in captivity and actively managed in breeding programs in which often little is known about the founder individuals. Recent developments in genetic research techniques have made it possible to sequence and study whole genomes. In this study we used the critically endangered Visayan warty pig (*Sus cebifrons*) as a case study to test the use of genomic information as a tool in conservation management. Two captive populations of *S. cebifrons* exist, which originated from two different Philippine islands. We found some evidence for a recent split between the two island populations; however all individuals that were sequenced show a similar demographic history. Evidence for both past and recent inbreeding indicated that the founders were at least to some extent related. Together with this, the low level of nucleotide diversity compared to other *Sus* species potentially poses a threat to the viability of the captive populations. In conclusion, genomic techniques answered some important questions about this critically endangered mammal and can be a valuable toolset to inform future conservation management in other species as well.

1. Introduction

The list of threatened and endangered species is growing rapidly, due to various anthropogenic causes. Current management of endangered species includes *in situ* and *ex situ* measurements. *In situ*, that is, within the range of the species, most conservation actions focus on habitat protection (protected areas), law enforcement (for reducing threats), and sometimes translocations or reintroductions [1]. For *ex situ* management, that is, outside the range of the species, such as in zoos or conservation centres, actions are mostly focused on keeping the population viable (both demographically and genetically) and as similar to the wild ancestor populations as possible (i.e., prevent adaptation to captivity [2]).

Despite the successes achieved with these approaches, there are also several challenges, for example, in prioritizing species for conservation. In trying to solve this problem, the International Union for the Conservation of Nature (IUCN) in 1994 initiated a scientific approach to categorise endangerment of species: the Red List of Threatened Species. This comprehensive list is currently a leading reference for governments, NGOs, and research institutions to decide on how to spend valuable resources for species conservation [1]. However, assessing a species properly is time-consuming and requires much information, which is often lacking. Additionally, not all species can be easily observed, creating a bias in the Red List towards the "easier observed" species [3]. If a species is not assessed well (or not assessed at all) the conservation actions planned for it might also miss their purpose. Another difficulty is the time that is needed to assess a species, understand its situation, implement suitable conservation actions, and wait for them to have an effect.

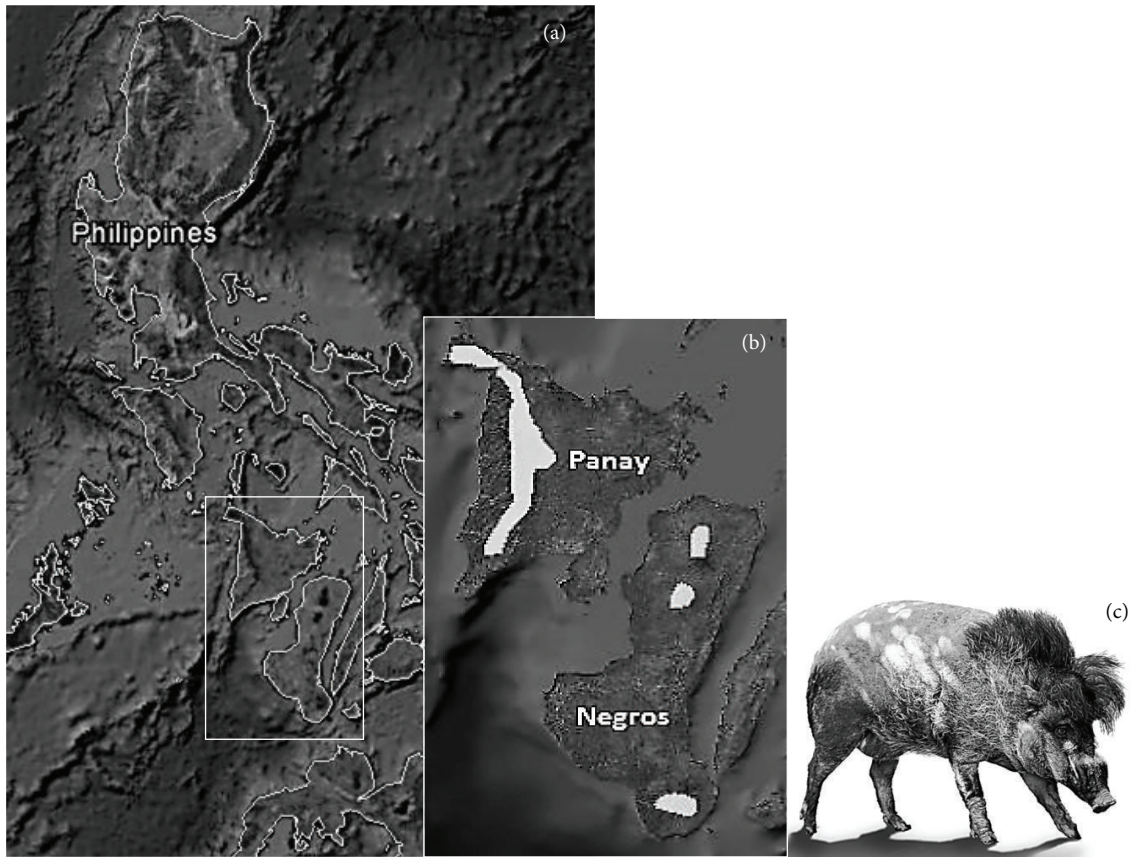


Figure 1: (a) Location of the islands Negros and Panay within the Philippine archipelago. At its smallest point the Guimaras Strait between them is 10 km wide. (b) Negros and Panay with the range of *Sus cebifrons* indicated in light-grey [4]. (c) *Sus cebifrons* individual (DVDW Photography).

With the current rate of extinction, this time might not always be available.

Nowadays, both as a legacy from earlier times and as a recent conservation measure, many threatened and endangered species are present in captivity [5]. These populations are often managed within a breeding program with the highest priority. This *ex-situ* management is mostly based on a pedigree, which requires information on relatedness between individuals. Especially for the individuals in the founder generation, however, information on relatedness is often missing. This causes breeding programs to make the potentially dangerous assumption that all founders are unrelated individuals [6]. If this assumption is not met, it can lead to unintended inbreeding events causing loss of genetic diversity and deleterious effects [7].

The development of genomic techniques has opened up many new opportunities for both breeding and population management and has caused a revolution in the field of commercial animal breeding [8, 9]. The recent developments make it possible to sequence and study the whole genome of individuals (genomics or next generation sequencing, NGS, [10]).

The main improvement of applying NGS, with respect to genetic methods such as microsatellites and SNP-arrays, is the enormous increase in loci that can be studied [11]. This allows for a more detailed study of previous research questions and opens up a completely new range of applications. For example, the possibility to study functional genetic variation (as opposed to neutral variation) interactions within the genome and interactions between the genome and the environment [12–15]. Although whole genome sequencing is still costly relative to targeted genotyping technologies, it rapidly becomes less expensive. Therefore it is expected that in the near future NGS techniques will be feasible for noncommercial research areas such as conservation biology as well [12, 16, 17].

The aim of this study is to apply currently available genomic techniques on whole genome sequencing data of a critically endangered mammal to test their application in species conservation and (captive) population management. It is expected that the outcomes of this study can be of direct use for conservation management of the species both *in situ* and *ex situ*.

As a case study, the critically endangered Visayan warty pig (*Sus cebifrons*) was used. *S. cebifrons* is endemic to the Philippines and mainly lives on the islands of Negros and Panay (10°00′N, 123°00′E and 11°09′N, 122°29′E, resp.) (Figure 1). On Negros it is estimated by experts in the field that only 200–500 individuals remain, while this estimate is 500–>1000 for Panay. The main threats to the populations are

Table 1: Sample codes of resequencing data available for seven *Sus cebifrons* individuals, two from Rotterdam zoo and five from San Diego zoo.

Rotterdam (Negros)	San Diego (Panay)
SCEB02M01	SCEB01F01
SCEB02M02	SCEBKb14130
	SCEBKb16637
	SCEBKb16508
	SCEBKb17528

habitat loss due to commercial logging operations and slash-and-burn farming, hunting for meat, negative reputation (crop damage), and genetic contamination via hybridization with free-ranging domestic and feral pigs (*Sus scrofa*) [4]. Despite several *in situ* conservation measures [4, 18] the population remained small with a decreasing trend. Therefore it was decided to create two captive populations in zoos (*ex situ*). San Diego zoo received a breeding group from a conservation centre on Panay and in 2004 another group of 8 founders (4 male, 4 female) was moved from conservation centres on Negros to Rotterdam zoo in the Netherlands to found a European captive population [4]. The Visayan warty pig fits our aim perfectly because of its level of endangerment, the availability of genomic data, and the uncertainty about the degree of divergence between the two island populations. In addition, because the captive populations were founded recently, it may be possible to draw conclusions on the situation of the wild population from the genomic information of captive individuals. The main questions assessed in this study concern the presence of a substructure between the two island populations, the genetic status of the species including the (historic) demography that explains this status, and the assumption of unrelatedness for the founder individuals of the captive population.

2. Methods

2.1. Generation of Data. Demographic information on the captive populations was gathered from studbook files kept by regional (i.e., EU or US) studbook coordinators and analysed in the software program PMx [19]. From these files the relationship between the sampled individuals was extracted.

DNA samples were previously sequenced [20, 21]. The data is deposited at the European Nucleotide Archive (ENA) under accession numbers PRJEB9326 and ERP001813 for *S. cebifrons* and other individuals, respectively (http://www.ebi.ac.uk/ena/). Seven complete whole genome sequences of *Sus cebifrons individuals* were available, of which two came from Rotterdam zoo and five from San Diego zoo (Table 1).

Resequencing data of *Sus cebifrons* individuals was aligned against the *Sus scrofa* reference genome (version 10.2 [22]) using the unique alignment option of MOSAIK aligner [23] and variants (Single Nucleotide Polymorphisms or SNPs) were called using SAMtools mpileup (version 0.1.19 [24]). Only variants with a read-depth between 0.5 and 2.0 times the average (i.e., between 5x and 20x) were selected and stored in variant call format (.vcf) using VCFtools (version 0.1.1 [25]). Unless otherwise stated, these "filtered variants" were used for all analyses.

2.2. Population Structure. To assess the phylogenetic relationship between individuals of *S. cebifrons* and between *S. cebifrons* and other species/populations within the genus *Sus*, a phylogenetic analysis was carried out. Variants were called for individuals using ANGSD (minimal mapping quality 30, minimal base quality 15, and SNP p value of $1e^{-6}$ [26]). Pairwise distances between individuals were calculated using PLINK 1.9 [27, 28], and hierarchical clustering was done by neighbour joining [29].

To obtain insight in whether or not the individuals from the two different islands were admixed, we used the software Admixture [30]. Only biallelic variant sites from the filtered variants were used as input file. We initially set K (i.e., the number of source populations) to 2 because the individuals came from two different islands, but tested for different K-values (1, 2, 3, 4, 5, 6, and 7) by calculating cross-validation errors.

2.3. Demographic History. To derive an estimate of the historic effective population size of the population and possibly gather evidence for a population substructure, a Pairwise Sequential Markovian Coalescence (PSMC) model was used [31]. This software uses the time to most recent common ancestor of a diploid genome (determined by looking at the density of heterozygotes) to estimate the effective population size (Ne) in the (distant) past. The individual whole genome consensus sequence, called by SAMtools, was used as an input for this analysis. We used a generation time of five years (in concordance with the studbook files that showed a generation time for the captive population of 4,5 years) and a default mutation rate of $2{,}5 \times 10^{-8}$ [22].

Demographic history of the individuals was studied by analysing regions of homozygosity (ROHs). ROHs can be informative for the level of inbreeding of a population; long ROHs are indicative for recent consanguineous matings while short(er) ROHs indicate more distant inbreeding as ROHs will break down over time due to recombination and mutation [32]. ROH abundance and length over time therefore depend on recombination and mutation rate [20]. To identify the ROHs present in an individual we used a sliding window approach with bins of 10.000 bp [20]. We filtered the variants for read-depth (0.5–2.0 times the average) to exclude sites with a low read-depth (low reliability) and sites with a very high read-depth (possible sequencing errors or copy number variants). A correction for missing sites was done, by scaling the number of identified SNPs up from number of covered sites to the total bin length. Bins with less than 1000 sites covered (<10% of total bin) were excluded from analyses. Sex chromosomes were also excluded from analyses as it is known that the recombination landscape of these chromosomes is different from the autosomes [20]. Adapted from Bosse et al. [20] we defined a ROH as a region of at least twenty consecutive bins with a number of SNPs per bin <0.33 of the genomic average. The average number of SNPs per bin (nucleotide diversity) outside ROHs was used as a measure of genetic diversity present in an individual. All these measures were

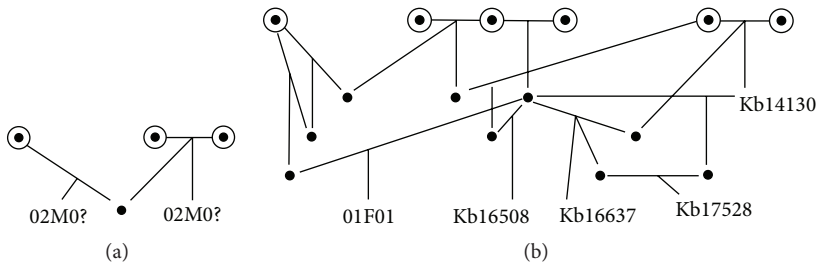


Figure 2: Part of the pedigree of the captive *Sus cebifrons* populations in Europe (a) and the US (b) reconstructed from the studbook files. Dots represent individuals, circles indicate founders (individuals with parents indicated as "WILD" in the studbook file), and sampled individuals are represented by their number as described in Table 1. For the two individuals from Rotterdam zoo, it is not known which sample is from which individual, and therefore they are indicated with a question mark.

also done for other *Sus* species for comparison (description of other species is in Supplementary Material, Table S1, available online at http://dx.doi.org/10.1155/2016/5613862).

Besides nucleotide diversity outside ROHs, all filtered variants were analysed with the Ensemble Variant Effect Predictor tool (VEP [33]). This tool was set to look for variants in coding regions only. Nonsynonymous variants were also annotated with SIFT (Sort Intolerant From Tolerant [34]) scores. SIFT scores range from 1 ("tolerated") to 0 ("deleterious"). A site is classified as deleterious when the variant in the genome leads to a different amino-acid in a protein, which in turn leads to the protein having different characteristics, for example, in shape and function. A SIFT score close to zero infers that the identified SNP is likely to have an effect, but the nature and extent of the effect cannot be ascertained. Therefore, the term "not-tolerated" (as opposed to "tolerated") will be used throughout this study to refer to sites with a SIFT score of (or close to) zero. For the VEP analysis and subsequent partitioning of variants between individuals and islands, variants were called again based on the positions identified previously, based on individually called genotypes. A multi-individual VCF was thus constructed using SAMtools mpileup (version 0.1.19 [24]). Only SNPs that were not fixed differences between *Sus scrofa* and *S. cebifrons* were retained. Furthermore, the minor allele count for variants to be considered was 2, to remove spurious allele calls as much as possible. Subsequently, variants were annotated using VEP based on Ensembl v83.

3. Results

3.1. Population Structure. The pedigree of the captive populations for both breeding programs was extracted from the studbook files (partly visualized in Figure 2). From this pedigree, the inbreeding coefficient (F) was calculated for the sampled individuals (Table 2). The phylogenetic analysis showed two clusters within *Sus cebifrons* (Figure 3), separating the two island populations. The time of the split between these two clades of *Sus cebifrons* is comparable to the split between the two different European wild boar populations, which was estimated at about 1 million years ago [22]. Additionally, the Admixture analysis, when forced to use a K-value of 2, also identified the two island populations.

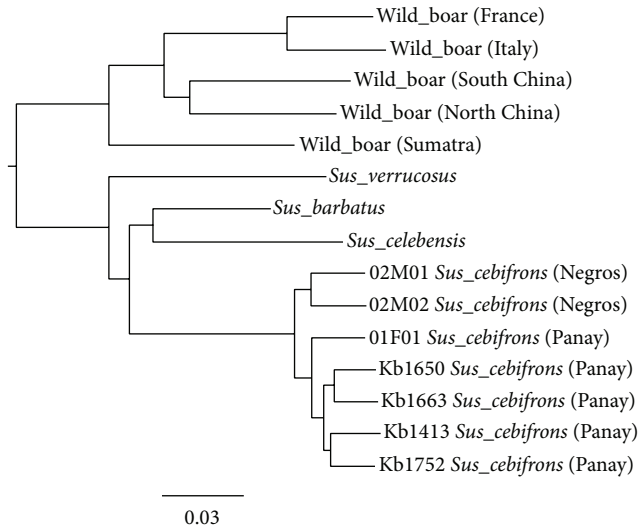


Figure 3: Neighbour joining, midpoint rooted phylogenetic tree of seven *Sus cebifrons* individuals, including other species for comparison (for description, see Table S1 in Supplementary Material). Wild boar samples represent different populations of *Sus scrofa*. Within the *Sus cebifrons* cluster, a split is visible, separating the two individuals from Negros (top) and the five individuals from Panay. The two smaller clusters within the Panay individuals are caused by relatedness between the sequenced individuals (Figure 2).

However, the cross-validation error was lowest for a K-value of 6 (Supplementary Material, Table S2). These results both indicate that a substructure is present, at least to some extent.

3.2. Demographic History. The PSMC analyses did not show a divergence between the two islands; all individuals showed a similar pattern of historic effective population size. As the PSMC shows the historic effective population size between 10.000 and 1.000.000 years ago; this indicates that the substructure found in the phylogenetic and Admixture analyses arose only recently. The PSMC results also show two severe bottlenecks in all populations, inferred from the individual genomes, one occurring around 100.000 years ago, and another more recent one which coincides with the end of the Last Glacial Maximum (LGM; Figure 4). Both bottlenecks

Table 2: Overview of number of ROHs, average length of ROHs, and nucleotide diversity outside ROHs for all *Sus cebifrons* individuals and other *Sus* individuals. Description of other *Sus* species is in Supplementary Material, Table S1. Asian wild boars 1 and 2 originate from north China and south China, respectively. European wild boars 1 and 2 represent populations from Italy and France, respectively.

Sample code or species name	Sex	Inbreeding coefficient (F)	# ROHs (>20 bins)	Average length of ROHs (kb)	Nucleotide diversity outside ROHs (π)
02M01	M	0	34	761.43	12.5
02M02	M	0	75	1764.21	12.1
01F01	F	0.1875	110	2245.32	11.9
Kb16508	M	0.0625	132	1420.68	12.4
Kb16637	F	0	144	1860.76	12.4
Kb17528	F	0.25	193	2867.06	12.0
Kb14130	F	0	130	2563.13	12.2
Sus verrucosus	U	NA	275	3829.13	6.3
Sus barbatus	U	NA	11	1551.82	26.4
Sus celebensis	U	NA	36	2719.72	23.2
Asian domestic	M	NA	271	2796.31	30.9
Asian wild boar 1	U	NA	155	1966.39	28.9
Asian wild boar 2	U	NA	44	2120.00	33.9
Asian wild boar (Japan)	U	NA	1172	1573.41	17.5
European domestic	F	NA	493	1859.01	28.4
European wild boar 1	M	NA	592	1912.82	16.2
European wild boar 2	U	NA	708	1401.60	18.4

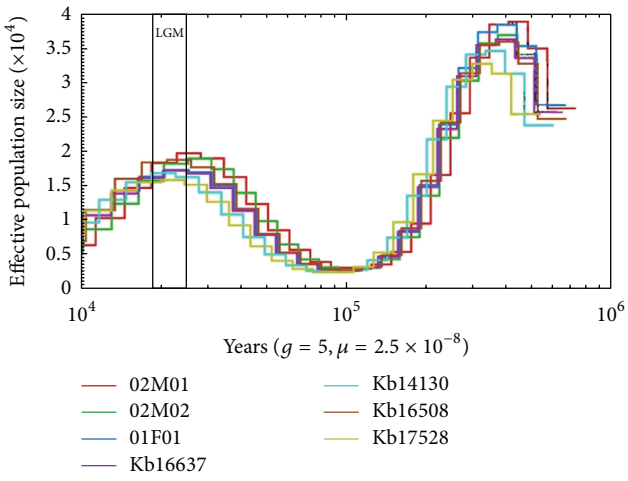


Figure 4: Estimated effective population size of *Sus cebifrons* based individual genomes, generation time (i.e., 5 years), and mutation rate ($2.5 * 10^{-8}$) from 10.000 (left) to 1.000.000 (right) years ago. The red and green lines indicate individuals from Rotterdam (originating from Negros island); the other individuals originated from San Diego (Panay island). A severe population bottleneck is found in all individuals around 100.000 years ago. The Last Glacial Maximum (LGM) occurred roughly 17.000 years ago (black rectangle).

are also present in most other *Sus* species (Supplementary Material, Figure S1 [35]).

For the demographic analyses we also identified regions of homozygosity (ROHs) and nucleotide diversity. On average we found 117±51 (average ± sd) ROHs in the *Sus cebifrons* individuals. This average was lower for Negros individuals than for Panay individuals, 55 ± 29 and 142 ± 31, respectively (Table 2). The average length of the identified ROHs was 1.9 ± 0.7 Mb. Here also a lower value for the Negros individuals was found, with 1.3 ± 0.7 Mb compared to 2.2 ± 0.6 for the Panay individuals. These numbers are similar as those found for other *Sus* species analysed using the same criteria (Table 2). In most *Sus cebifrons* individuals, the largest proportion of the genome was covered by ROHs in the longest category (Supplementary Material, Figure S2). However, most ROHs in all *Sus cebifrons* individuals fell within the shortest length category of 0.2–0.5 Mb. Although logic predicts the longest category would cover the largest proportion of the genome, a high number of short ROHs could easily cover a proportion of the genome larger than a few long ROHs. This is also shown by individuals of other *Sus* species (Supplementary Material, Figure S3). The large proportion of coverage by long ROHS is an indication of recent inbreeding [20].

Nucleotide diversity outside the ROHs did not differ between the islands and was on average 12 SNPs per bin (10 kb) in all *Sus cebifrons* individuals (Table 2) and seemed to follow a normal distribution (Supplementary Material, Figure S4). The other individuals showed on average 23 SNPs per bin (Table 2). The very low nucleotide diversity outside the ROHs is probably a direct effect of the extreme bottleneck found in the PSMC analysis, as a bottleneck generally causes rapid loss of genetic variation.

For variant effect prediction, 4679012 variants were retained. Of these, 38321 were exonic, 11532 were non-synonymous but classified as tolerated according to SIFT predictions, and only 3884 were predicted to be not-tolerated

Table 3: Assessment of shared and island-specific variation. The vast majority of the SNPs, both coding and noncoding, are shared between islands, and only a very small portion may be specific. Note that this pertains a total of 4679012 SNPs, excluding fixed differences between *S. scrofa* and *S. cebifrons*, for which the minor allele count is at least two out of seven individuals, or 14 haplotypes.

	Shared	Panay ($N = 5$)	Negros ($N = 2$)
All (4679012 SNPs total)	3969361	457741	251910
Synonymous	19565	2150	1190
Nonsynonymous tolerated	10081	918	533
Nonsynonymous not-tolerated	3370	343	171

(Table 3). Only a fraction—around 15%—of the variants was specific to either one of the islands based on the small population sample surveyed here (Supplementary Material, Table S3).

4. Discussion

The genomic analyses showed that, at least to some extent, a substructure is present between the two island populations of *Sus cebifrons*. This was visible in the results of the Admixture analysis and in the presence of island-specific variation. However, the PSMC analysis showed a similar demographic history for all individuals, regardless of their source population, suggesting one population of origin for the sampled individuals. This indicates that the present structure only arose recently in evolutionary terms (as the PSMC analysis provides estimates for 10.000–1.000.000 years ago). This hypothesis is supported by the relatively recent split in the phylogeny, suggesting that the split between European wild boar populations predated the split between the *Sus cebifrons* populations on both Philippine islands. For comparison, the split between European and Asian wild boar took place about 1 million years ago but the populations are considered one species [22]. Further evidence for a very recent split between the populations is the large proportion of shared variants between islands and thus the small fraction of island-specific variation.

The extremely low nucleotide diversity found in the analysed individuals, as compared to other species, is probably a result of the bottleneck visible in the PSMC result. The number of short ROHs present in the genomes is indicative of past inbreeding. The data also showed long ROHs that are signs of inbreeding in recent generations. Because both captive populations in the EU and US are under strict management to minimize inbreeding, it is plausible that some relatedness was already present in the founders of the two captive populations. The small population sizes present at the islands are indicative of this as well. In addition, it is not clear whether the founders were a representative sample of the island populations. It is clear from our results that the assumption of founder-unrelatedness was violated.

In the captive populations in the US and EU, numbers have increased rapidly since the founder generation. However, because reproduction was not equally successful for all individuals, the amount of genetic diversity present now is not the maximum that could have been retained. Furthermore, the extremely low nucleotide diversity and the signs of recent inbreeding (long ROHs) found in the current generation potentially threaten the viability of the captive populations. If populations are to be kept separately in the future, inbreeding in each of the populations has to be limited. However, with regard to the problems mentioned above, based on the similarity in demographic history of the individuals, it might be prudent to decide to merge the two breeding programs in order to increase the viability of the total captive population and the probability of reintroduction. By doing this the genetic diversity, and with that the potential for adaptation, will increase. Moreover, given the extremely shallow genetic divergence between the islands, problems of outbreeding depression are not expected.

A decision for merging captive populations cannot be based on the present study alone. Further research should focus on deleterious load present in both captive populations. Purging in these naturally small populations may have removed variants that are deleterious in homozygous state [36]. The ratio between heterozygous and homozygous states in the not-tolerated variants found in this study is an indication that purging has removed some variants from the populations (Supplementary Material, Table S4). Merging the two captive populations would increase the frequency of these variants, which could have deleterious effects. The decision to merge two captive populations should therefore be made with caution. Information on, for example, the heterozygous/homozygous ratio of variants can help in making informed decisions.

Although it is not the focus of this study, the same genomic methods as described here can be used to select individuals for breeding [21]. Individuals can be selected based on deleterious load or identity-by-descent (IBD) segments. It has been shown that inbreeding measures based on ROHs are more reliable than inbreeding estimates from a pedigree [32]. Also, simulations showed that management based on molecular ancestry and IBD segments resulted in higher maintained genetic diversity and fitness as opposed to management based on a pedigree [21]. In the IBD segments management scenario the length of the segments was crucial for this result: in longer IBD segments there is a higher chance of homozygous deleterious alleles as these regions have a common ancestor [21]. By using genomic information in a breeding program, negative effects from inbreeding and deleterious load can be more actively avoided.

The results of the genomic analyses as presented in this study show that they can be of direct use for conservation management either *in situ* or *ex situ*, even with the small sample sizes generally available in conservation settings. *In situ* the identification of a substructure can lead to

reassessing priorities for conservation. Identifying hybrids (as hybridization is a big threat to wildlife worldwide [37]) can give better insights into the effect of the threat and can help select individuals for (captive) breeding programs. In addition, an analysis of historic population size can explain levels of nucleotide diversity present in the population and put the current numbers in a historic perspective. *Ex situ,* the selection of founder individuals [38] and identification of relatedness between founder individuals can lead to more informed reproductive planning, resulting in higher levels of genetic diversity maintained. This can even be increased by assessing individual inbreeding levels or identifying carriers of deleterious alleles and incorporating this information in breeding recommendations [39]. To do this for all individuals in the current (captive) breeding programs is, for the time being, very time-consuming and costly. However, knowing this information from the founders, as well as being able to monitor and model the behaviour of the genetic material through the breeding program, is feasible. For example, if genomic information is known for the founders, genetic markers such as a subset of SNPS or microsatellites uniquely identifying the founders should suffice to adequately manage the next generations for inbreeding and deleterious variation. This combination of techniques is cost-efficient but does require information from the founders, which is not always available.

Another way of using genomic information without sequencing all individuals is to model the population and simulate what will happen with the genetic material under certain management strategies. There are already software programs available that can do just this. Examples are the PMx 2000 software [19], which monitors a population and can project future demographics under certain management strategies, and Vortex software [40] that uses demographic and stochastic factors to calculate an extinction risk of a population (a Population Viability Analysis, PVA). Some genetic information, such as inbreeding (modelled as a default value for deleterious load and the effect on juvenile mortality), can already be incorporated in these tools and thus be modelled over time. This information however is not species-specific and far from the detailed level of information that is available with genomic analyses [41]. New insights gained by genomic studies can provide more detailed input for these or similar software programs and can be used in concordance with other data sources [42, 43]. Incorporation of this information in management tools has been identified as a "conservation priority" [41] and can lead to less uncertainty and more successful breeding programs.

An example of a measure that could be of interest for these modelling efforts is the number and length of ROHs. It has already been found that the selection of individuals for participating in breeding programs based on ROHs gave the best results in maintaining diversity without losing much fitness, as compared to optimal contributions and including inbred matings (purging of deleterious variation) [44].

Although genomic data can already provide extremely useful information for the development of (conservation) management strategies, we need to understand more about the genomic measures and what they actually tell us about a population before they can be applied on a large scale. For example, about ROHs: how does the recombination landscape of the genome affect the distribution of ROHs over time? And how exactly does ROH-breakdown affect nucleotide diversity? Are these characteristics species-specific or is it possible to derive some general "rules-of-thumb"? With these uncertainties in mind, the incorporation of genomic analyses in conservation management may seem to be something of the distant future. Not least because the translation of academic knowledge to conservation practice is often slow [45]. However, the rate of development of sequencing technology has progressed very rapidly over the past decade and is expected to continue to do so, potentially enabling extremely cheap, whole genome variation information for all actively managed populations in the near future. Even today, it is already possible to use genomic analyses for individual cases of critically endangered species such as the Visayan warty pig in this example. Other examples include the California condor [46–48] and the North-American bison [49].

5. Conclusion

Genomic techniques represent a promising new toolset in the field of conservation biology. In this study genomic data analyses answered several questions regarding the captive population of the critically endangered Visayan warty pig. We found evidence for a recent split between the two island populations. However, with the current level of inbreeding, the viability of the total captive population and the probability of reintroduction might increase by merging the two captive populations. With the current rate of development, and associated lower costs, it is expected that genomic techniques will be feasible for broad application in conservation biology in the near future.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors thank Laurent Frantz and Ronald Bontrop for valuable discussions about the methodology and interpretation of results and Harald Schmidt, Janno Weerman, Aryanna Herscheid, and Angela Glatston for collaboration. Samples were provided by Rotterdam zoo, Netherlands, and San Diego zoo, US. They thank DVDW Photography for kindly providing the picture in Figure 1(c).

References

[1] IUCN 2013, "The IUCN Red List of Threatened Species," Version 2013.1., http://www.iucnredlist.org.

[2] R. Frankham, "Genetic adaptation to captivity in species conservation programs," *Molecular Ecology*, vol. 17, no. 1, pp. 325–333, 2008.

[3] A. S. L. Rodrigues, J. D. Pilgrim, J. F. Lamoreux, M. Hoffmann, and T. M. Brooks, "The value of the IUCN Red List for conservation," *Trends in Ecology and Evolution*, vol. 21, no. 2, pp. 71–76, 2006.

[4] W. Oliver, "Sus cebifrons," in *IUCN 2013. IUCN Red List of Threatened Species*, Version 2013.1., 2008, http://www.iucnredlist.org/.

[5] D. A. Conde, N. Flesness, F. Colchero, O. R. Jones, and A. Scheuerlein, "An emerging role of zoos to conserve biodiversity," *Science*, vol. 331, no. 6023, pp. 1390–1391, 2011.

[6] J. A. Rudnick and R. C. Lacy, "The impact of assumptions about founder relationships on the effectiveness of captive breeding strategies," *Conservation Genetics*, vol. 9, no. 6, pp. 1439–1450, 2008.

[7] L. F. Keller and D. M. Waller, "Inbreeding effects in wild populations," *Trends in Ecology and Evolution*, vol. 17, no. 5, pp. 230–241, 2002.

[8] N. D. Beuzen, M. J. Stear, and K. C. Chang, "Molecular markers and their use in animal breeding," *The Veterinary Journal*, vol. 160, no. 1, pp. 42–52, 2000.

[9] B. J. Hayes, P. J. Bownan, A. J. Chamberlain, and M. E. Goddard, "Invited review: genomic selection in dairy cattle: progress and challenges," *Journal of Dairy Science*, vol. 92, no. 2, pp. 433–443, 2009.

[10] J. E. Pool, I. Hellmann, J. D. Jensen, and R. Nielsen, "Population genetic inference from genomic sequence variation," *Genome Research*, vol. 20, no. 3, pp. 291–300, 2010.

[11] F. W. Allendorf, P. A. Hohenlohe, and G. Luikart, "Genomics and the future of conservation genetics," *Nature Reviews Genetics*, vol. 11, no. 10, pp. 697–709, 2010.

[12] M. H. Kohn, W. J. Murphy, E. A. Ostrander, and R. K. Wayne, "Genomics and conservation genetics," *Trends in Ecology and Evolution*, vol. 21, no. 11, pp. 629–637, 2006.

[13] N. J. Ouborg, C. Pertoldi, V. Loeschcke, R. K. Bijlsma, and P. W. Hedrick, "Conservation genetics in transition to conservation genomics," *Trends in Genetics*, vol. 26, no. 4, pp. 177–187, 2010.

[14] N. J. Ouborg, F. Angeloni, and P. Vergeer, "An essay on the necessity and feasibility of conservation genomics," *Conservation Genetics*, vol. 11, no. 2, pp. 643–653, 2010.

[15] F. Angeloni, N. Wagemaker, P. Vergeer, and J. Ouborg, "Genomic toolboxes for conservation biologists," *Evolutionary Applications*, vol. 5, no. 2, pp. 130–143, 2012.

[16] O. A. Ryder, "Conservation genomics: applying whole genome studies to species conservation efforts," *Cytogenetic and Genome Research*, vol. 108, no. 1–3, pp. 6–15, 2005.

[17] C. C. Steiner, A. S. Putnam, P. E. A. Hoeck, and O. A. Ryder, "Conservation genomics of threatened animal species," *Annual Review of Animal Biosciences*, vol. 1, pp. 261–281, 2013.

[18] R. A. Blouch, "Conservation and research priorities for threatened suids of south and southeast Asia," *IBEX Journal of Mountain Ecology*, vol. 3, pp. 21–25, 1995.

[19] J. D. Ballou, R. C. Lacy, and J. P. Pollak, *PMx: Software for Demographic and Genetic Analysis and Management of Pedigreed Populations*, Chicago Zoological Society, Brookfield, Ill, USA, 2010.

[20] M. Bosse, H.-J. Megens, O. Madsen et al., "Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape," *PLoS Genetics*, vol. 8, no. 11, Article ID e1003100, 2012.

[21] M. Bosse, H. Megens, O. Madsen et al., "Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations," *Genome Research*, vol. 25, no. 7, pp. 970–981, 2015.

[22] M. A. Groenen, A. L. Archibald, H. Uenishi et al., "Analyses of pig genomes provide insight into porcine demography and evolution," *Nature*, vol. 491, no. 7424, pp. 393–398, 2012.

[23] W. P. Lee, M. P. Stromberg, A. Ward, C. Stewart, E. P. Garrison, and G. B. Marth, "MOSAIK: a harsh-based algorithm for accurate next-generation sequencing short read mapping," *PLoS ONE*, vol. 9, no. 3, Article ID e90581, 2014.

[24] H. Li, B. Handsaker, A. Wysoker et al., "The sequence alignment/map format and SAMtools," *Bioinformatics*, vol. 25, no. 16, pp. 2078–2079, 2009.

[25] P. Danecek, A. Auton, G. Abecasis et al., "The variant call format and VCFtools," *Bioinformatics*, vol. 27, no. 15, Article ID btr330, pp. 2156–2158, 2011.

[26] T. S. Korneliussen, A. Albrechtsen, and R. Nielsen, "ANGSD: analysis of next generation sequencing data," *BMC Bioinformatics*, vol. 15, no. 1, article 356, 2014.

[27] S. Purcell, B. Neale, K. Todd-Brown et al., "PLINK: a tool set for whole-genome association and population-based linkage analyses," *The American Journal of Human Genetics*, vol. 81, no. 3, pp. 559–575, 2007.

[28] C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee, "Second-generation PLINK: rising to the challenge of larger and richer datasets," *GigaScience*, vol. 4, article 7, 2015.

[29] Phylip, 2013, http://evolution.genetics.washington.edu/phylip.html.

[30] D. H. Alexander, J. Novembre, and K. Lange, "Fast model-based estimation of ancestry in unrelated individuals," *Genome Research*, vol. 19, no. 9, pp. 1655–1664, 2009.

[31] H. Li and R. Durbin, "Inference of human population history from individual whole-genome sequences," *Nature*, vol. 475, no. 7357, pp. 493–496, 2011.

[32] I. Curik, M. Ferenčaković, and J. Sölkner, "Inbreeding and runs of homozygosity: a possible solution to an old problem," *Livestock Science*, vol. 166, no. 1, pp. 26–34, 2014.

[33] W. McLaren, B. Pritchard, D. Rios, Y. Chen, P. Flicek, and F. Cunningham, "Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor," *Bioinformatics*, vol. 26, no. 16, pp. 2069–2070, 2010.

[34] P. C. Ng and S. Henikoff, "SIFT: predicting amino acid changes that affect protein function," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3812–3814, 2003.

[35] L. A. F. Frantz, J. G. Schraiber, O. Madsen et al., "Genome sequencing reveals fine scale diversification and reticulation history during speciation in *Sus*," *Genome Biology*, vol. 14, no. 9, article R107, 2013.

[36] S. Glémin, "How are deleterious mutations purged? Drift versus nonrandom mating," *Evolution*, vol. 57, no. 12, pp. 2678–2687, 2003.

[37] D. Simberloff, "Hybridization between native and introduced wildlife species: importance for conservation," *Wildlife Biology*, vol. 2, no. 3, pp. 143–150, 1996.

[38] W. Miller, S. J. Wright, Y. Zhang, S. C. Schuster, and V. M. Hayes, "Optimization methods for selecting founder individuals for captive breeding or reintroduction of endangered species," in *Proceedings of the 15th Pacific Symposium on Biocomputing (PSB '10)*, pp. 43–53, January 2010.

[39] S. Fritz, A. Capitan, A. Djari et al., "Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in GART, SHBG and SLC37A2," *PLoS ONE*, vol. 8, no. 6, Article ID e65550, 2013.

[40] R. C. Lacy and J. P. Pollak, *Vortex: A Stochastic Simulation of the Extinction Process. Version 10.0*, Chicago Zoological Society, Brookfield, Ill, USA, 2013.

[41] R. Frankham, "Where are we in conservation genetics and where do we need to go?" *Conservation Genetics*, vol. 11, no. 2, pp. 661–663, 2010.

[42] E. S. Fienieg and P. Galbusera, "The use and integration of molecular DNA information in conservation breeding programmes: a review," *Journal of Zoo and Aquarium Research*, vol. 1, no. 2, pp. 44–51, 2013.

[43] M. T. Olsen, L. W. Andersen, R. Dietz, J. Teilmann, T. Härkönen, and H. R. Siegismund, "Integrating genetic data and population viability analyses for the identification of harbour seal (*Phoca vitulina*) populations and management units," *Molecular Ecology*, vol. 23, no. 4, pp. 815–831, 2014.

[44] M. Á. R. De Cara, B. Villanueva, M. Á. Toro, and J. Fernández, "Using genomic tools to maintain diversity and fitness in conservation programmes," *Molecular Ecology*, vol. 22, no. 24, pp. 6091–6099, 2013.

[45] A. B. A. Shafer, J. B. W. Wolf, P. C. Alves et al., "Genomics and the challenging translation into conservation practice," *Trends in Ecology and Evolution*, vol. 30, no. 2, pp. 78–87, 2015.

[46] M. N. Romanov, E. M. Tuttle, M. L. Houck et al. et al., "The value of avian genomics to the conservation of wildlife," *BMC Genomics*, vol. 10, supplement 2, p. S10, 2009.

[47] K. Ralls, J. D. Ballou, B. A. Rideout, and R. Frankham, "Genetic management of chondrodystrophy in California condors," *Animal Conservation*, vol. 3, no. 2, pp. 145–153, 2000.

[48] K. Ralls and J. D. Ballou, "Genetic status and management of California condors," *The Condor*, vol. 106, no. 2, pp. 215–228, 2004.

[49] N. D. Halbert, T. J. Ward, R. D. Schnabel, J. F. Taylor, and J. N. Derr, "Conservation genomics: disequilibrium mapping of domestic cattle chromosomal segments in North American bison populations," *Molecular Ecology*, vol. 14, no. 8, pp. 2343–2362, 2005.