Conservation [and] genomics of free-ranging populations

Aaron Shafer
aarons.shafer@trentu.ca
@shaferab

01 June 2016 - IUFRO Genomics and Forest Tree Genetics
Why I am standing up here

Provocative workshop and paper

ConGenOmics workshop
Wiks Slott, Uppsala
18th – 20th of March 2014

Academic exercise or transition with real-world implications

Genomics and the challenging translation into conservation practice

Aaron B.A. Shafer¹, Jochen B.W. Wolf², Paulo C. Alves³, Linnea Bergström⁴, Michael W. Bruford⁵, Ioana Brännström⁶, Guy Colling⁷, Love Dalén⁸, Luc De Meester⁹, Robert Ekblom¹⁰, Katie D. Fawcett¹¹, Simone Fior¹², Mehrdad Hajibabaei¹³, Jason A. Hill¹⁴, A. Rus Hözel¹⁵, Jacob Höglund¹⁶, Evelyn L. Jensen¹⁷, Johannes Krause¹⁸, Torsten N. Kristensen¹⁹, Michael Krützen²⁰, John K. McKay²¹, Anita J. Norman²², Rob Ogden²³, E. Martin Österling²⁴, N. Joop Ouborg²⁵, John Piccolo²⁶, Danijela Popović²⁷, Craig R. Primmer²⁸, Floyd A. Reed²⁹, Marie Roumet³⁰, Jordi Salmona³¹, Tamara Schenekar³², Michael K. Schwartz³³, Gernot Segelbacher³⁴, Helen Senn³⁵, Jens Thaulow³⁶, Mia Valtonen³⁷, Andrew Veale³⁸, Philippe Vergeer³⁹, Nagarjun Vijay⁴⁰, Carles Vilà⁴¹, Matthias Weissensteiner⁴², Lovisa Wennerström⁴³, Christopher W. Wheat⁴⁴, and Piotr Zielinski⁴⁵

¹Department of Biological Sciences, University of Alberta, Canada
²Laboratory of Evolutionary Genetics, Institute of Evolutionary Biology, University of Edinburgh, UK
³Department of Language and Culture, School of Psychology and Education, University of the State of Porto, Portugal
⁴Department of Evolutionary Biology, Uppsala University, Sweden
⁵Department of Anthropology, Columbia University, New York, USA
⁶Department of Natural History, University of Bergen, Norway
⁷Department of evolutionary Ecology and Conservation, University of Leiden, The Netherlands
⁸Department of Evolutionary Biology, University of Gothenburg, Sweden
⁹Department of Biology, University of Southern Denmark, Denmark
¹⁰Department of Zoology, University of Oxford, UK
¹¹Department of Biological Sciences, University of Alberta, Canada
¹²Department of Evolutionary Biology, Uppsala University, Sweden
¹³Department of Biology, University of Southern Denmark, Denmark
¹⁴Department of Biological Sciences, University of Alberta, Canada
¹⁵Department of Evolutionary Genetics, Institute of Evolutionary Biology, University of Edinburgh, UK
¹⁶Department of Evolutionary Biology, Uppsala University, Sweden
¹⁷Department of Evolutionary Biology, Uppsala University, Sweden
¹⁸Musée National d'Histoire Naturelle, France
¹⁹Department of Environmental Studies, University of Helsinki, Finland
²⁰Centre for Integrative Biology, University of Porto, Portugal
²¹Department of Biological Sciences, University of Alberta, Canada
²²Laboratory of Evolutionary Genetics, Institute of Evolutionary Biology, University of Edinburgh, UK
²³Department of Anthropology, Columbia University, New York, USA
²⁴Department of Natural History, University of Bergen, Norway
²⁵Department of Evolutionary Biology, University of Leiden, The Netherlands
²⁶Department of Evolutionary Biology, University of Gothenburg, Sweden
²⁷Department of Biological Sciences, University of Alberta, Canada
²⁸Department of Evolutionary Genetics, Institute of Evolutionary Biology, University of Edinburgh, UK
²⁹Department of Biology, University of Southern Denmark, Denmark
³⁰Department of Zoology, University of Oxford, UK
³¹Department of evolutionery Ecology and Conservation, University of Leiden, The Netherlands
³²Department of Natural History, University of Bergen, Norway
³³Department of Evolutionary Biology, University of Leiden, The Netherlands
³⁴Department of Evolutionary Biology, Uppsala University, Sweden
³⁵Department of Zoology, University of Oxford, UK
³⁶Department of Evolutionary Biology, University of Gothenburg, Sweden
³⁷Department of Biology, University of Southern Denmark, Denmark
³⁸Department of Zoology, University of Oxford, UK
³⁹Department of Evolutionary Genetics, Institute of Evolutionary Biology, University of Edinburgh, UK
⁴⁰Centre for Integrative Biology, University of Porto, Portugal
⁴¹Department of Biology, University of Southern Denmark, Denmark
⁴²Department of Biological Sciences, University of Alberta, Canada
⁴³Department of Evolutionary Biology, Uppsala University, Sweden
⁴⁴Department of Evolutionary Biology, Uppsala University, Sweden
⁴⁵Department of Zoology, University of Oxford, UK
Conservation genetics has a relatively long history
Bonnell & Selander 1976 (Science)

Elephant Seals: Genetic Variation and Near Extinction

Abstract. Blood samples from northern elephant seals (Mirounga angustirostris), representing five breeding colonies in California and Mexico, were surveyed electrophoretically for protein variation reflecting underlying genetic differences. No polymorphisms were found among 21 proteins encoded by 24 loci. This uniform homozygosity may be a consequence of fixation of alleles brought about by the decimation of this species by sealers in the last century.

Michael L. Bonnett
Thimann Laboratories, University of California, Santa Cruz 95060
Robert K. Selander
Department of Zoology, University of Texas, Austin 78712

animal species. Our results also suggest that the northern elephant seal, now lacking a pool of variability with which to adapt to changing conditions, is especially vulnerable to environmental modification.
Strategic Goal C: To improve the status of biodiversity by safeguarding ecosystems, species and genetic diversity.

Distinct Population Segment: quantitative genetic separation.... differ in genetic characteristics...
Early 2000’s genomics enters the conservation discussion
All selected articles predate 2010
Observations about conservation genetic studies

From Baillie et al. (2010)
Conservation genetics problems have genomic solutions!
Many accepted data streams and applications

From Allendorf et al. (2010)
**Conservation genetics problems have genomic solutions!**

Many accepted data streams and applications

<table>
<thead>
<tr>
<th>Primary problem</th>
<th>Possible genomic solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation of $N_e$, $m$ and $s$</td>
<td>Increasing the number of markers, reconstructing pedigrees and using haplotype information will provide greater power to estimate and monitor $N_e$ and $m$, as well as to identify migrants, estimate the direction of migration and estimate $s$ for individual loci within a population.</td>
</tr>
<tr>
<td>Reducing the amount of admixture in hybrid populations</td>
<td>Genome scanning of many markers will help to identify individuals with greater amounts of admixture so that they can be removed from the breeding pool.</td>
</tr>
<tr>
<td>Identification of units of conservation: species, evolutionarily significant units and management units</td>
<td>The incorporation of adaptive genes and gene expression will augment our understanding of conservation units based on neutral genes. The use of individual-based landscape genetics will help to identify boundaries between conservation units more precisely.</td>
</tr>
<tr>
<td>Minimizing adaptation to captivity</td>
<td>Numerous markers throughout the genome could be monitored to detect whether populations are becoming adapted to captivity.</td>
</tr>
<tr>
<td>Predicting harmful effects of inbreeding depression</td>
<td>Understanding the genetic basis of inbreeding depression will facilitate the prediction of the effectiveness of purging. Genotyping of individuals at loci associated with inbreeding depression will allow the selection of individuals as founders or mates in captive populations. Pedigree reconstruction will allow more powerful tests of inbreeding depression.</td>
</tr>
<tr>
<td>Predicting the intensity of outbreeding depression</td>
<td>Understanding the divergence of populations at adaptive genes will help to predict effects on fitness when these genes are combined. Detecting chromosomal rearrangements will help to predict outbreeding depression.</td>
</tr>
<tr>
<td>Predicting the viability of local populations</td>
<td>Incorporating genotypes that affect vital rates and the genetic architecture of inbreeding depression will improve population viability models.</td>
</tr>
<tr>
<td>Predicting the ability of populations to adapt to climate change and other anthropogenic challenges</td>
<td>Understanding adaptive genetic variation will help to predict the response to a rapidly changing environment or to harvesting by humans and allow the selection of individuals for assisted migration.</td>
</tr>
</tbody>
</table>

From Allendorf et al. (2010)
What do the blogs are saying? @shaferab

In short: variant analysis is not a solved problem and in fact we are still working out basic standards… generally meet the “good enough” threshold.

Sequencing tech today is still rocky… The software is buggy… The datasets are dirty… do not have the scientific training to understand.

It would be responsible, however, for researchers to temper their hype — though this seems unlikely, because hype pays.
A beginners guide to SNP calling from high-throughput DNA-sequencing data

André Altmann · Peter Weber · Daniel Bader · Michael Preuß · Elisabeth B. Binder · Bertram Müller-Myhok

Abstract High-throughput DNA sequencing (HTS) is of increasing importance in the life sciences. One of its most prominent applications is the sequencing of whole genomes or targeted regions of the genome such as all exonic regions (i.e., the exome). Here, the objective is the identification of genetic variants such as single nucleotide polymorphisms (SNPs). The extraction of SNPs from the raw genetic sequences involves many processing steps and the application of a diverse set of tools. We review the essential building blocks for a pipeline that calls SNPs from raw HTS data. The pipeline includes quality control, mapping of short reads to the reference genome, visualization and post-processing of the alignment including base quality recalibration. The final steps of the pipeline include the SNP calling procedure along with filtering of SNP candidates. The steps of this pipeline are accompanied by an analysis of a publicly available whole-exome sequencing dataset. To this end, we employ several alignment programs and SNP calling routines for highlighting the fact that the choice of the tools significantly affects the final results.

Introduction

The initial sequencing of the entire human genome with its first draft published in 2001 was an effort that could only be accomplished by large research consortia, and still required a decade of time and large financial resources (Consortium 2004; Lander et al. 2001; Venter et al. 2001). The resulting blueprint of the human genome facilitated a number of follow-up technologies such as (in their current...
The impact of pipelines: site-frequency spectra from our sea lion data

94 individuals with ddRAD data and SNPs with various pipelines and filters

Hypothetical SFS
The impact of pipelines: site-frequency spectra from our sea lion data

94 individuals with ddRAD data and SNPs with various pipelines and filters

- Reference vs denovo
- Missing data
- Basic filters
The impact of pipelines: site-frequency spectra from our sea lion data

94 individuals with ddRAD data and SNPs with various pipelines and filters

- Expansion
- Bottleneck

- Reference vs denovo
- Missing data
- Basic filters

No. of chromosomes
The impact of pipelines: an example from our sea lion pedigree

7 trios (known mother - father - offspring)
Demographic inferences: examples from simulated data

1. Simulate genomic data (known scenarios)
2. Use population genomic methods (Approx. Bayesian Computation)
3. Compare estimated to true values
Table 4 Summary of the posterior sample for each parameter in model 3

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Mean</th>
<th>Difference</th>
<th>95% interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>0.1</td>
<td>0.1000</td>
<td>0.00%</td>
<td>[0.0654, 0.1449]</td>
</tr>
<tr>
<td>$m_{12}$</td>
<td>2</td>
<td>2.3454</td>
<td>17.27%</td>
<td>[0.4864, 4.6692]</td>
</tr>
<tr>
<td>$m_{21}$</td>
<td>1</td>
<td>0.8356</td>
<td>16.44%</td>
<td>[0.0686, 3.1288]</td>
</tr>
<tr>
<td>$N_1$</td>
<td>2,000</td>
<td>1,987</td>
<td>0.65%</td>
<td>[1,106, 3,178]</td>
</tr>
<tr>
<td>$N_2$</td>
<td>5,000</td>
<td>4,943</td>
<td>1.14%</td>
<td>[3,838, 5,828]</td>
</tr>
<tr>
<td>$N_{12}$</td>
<td>100,000</td>
<td>131,615</td>
<td>31.62%</td>
<td>[66,355, 192,624]</td>
</tr>
<tr>
<td>$N_4$</td>
<td>150,000</td>
<td>160,328</td>
<td>6.89%</td>
<td>[122,769, 194,806]</td>
</tr>
</tbody>
</table>

The true value, the mean of the posterior sample, the difference of the estimated mean value and the true value, and the 95% credible interval of the posterior sample are given for each parameter.
**RADseq data**
Shafer et al. (2015) in Molecular Ecology

---

**Model fit**
- **Prediction error**
  - $T_s$
  - $N_e$
  - $T_x$

**No. of Loci**
- 500
- 1000
- 10000
- 50000

**Density**
- $N_1$
- $N_2$
- $N_3$
- $N_4$

---

500 loci
1000 loci
10000 loci
50000 loci
Local adaptation
McMahon et al. (2014) in Evolutionary Applications

As argued above, we think that ‘local adaptation’ is the most important issue where genomics can contribute to conservation science. We want to stress that we do not see ‘local adaptation’ as different from the issue of ‘preserving genetic variation’ or ‘identifying ecotypes’. These aspects are instead tightly linked. Without genetic variation, there can be no local adaptation, and without local adaptation, no ecotypes. Further, simply because local adaptation is the most important aspect, this does not exclude, for example ‘estimation of demographic parameters’. Our argument is simply that the first is more important, not that the second is unimportant.

Conclusions and perspectives
We predict that genomics will make a difference primarily in determining which parts of the genomes are responsible for local adaptation and therefore important to preserve.
Detecting genetic basis to phenotypes requires A LOT of data
Kardos et al. (2016) in Molecular Ecology

Genome wide associations
Inbreeding estimates and genomic data
Hoffman et al. (2014) in Proc. Nat. Acad. Sci; Shafer et al. (In press) J. of Heredity

Heterozygosity fitness

Runs of Homozygosity

>>100,000 SNPs required
Conservation genetics problems have genomic solutions *that we’re still working on*

<table>
<thead>
<tr>
<th>Primary problem</th>
<th>Possible genomic solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation of $N_e$, $m$ and $s$</td>
<td>Increasing the number of markers, reconstructing pedigrees and using haplotype information will provide greater power to estimate and monitor $N_e$ and $m$, as well as to identify migrants, estimate the direction of migration and estimate $s$ for individual loci within a population</td>
</tr>
<tr>
<td>Reducing the amount of admixture in hybrid populations</td>
<td>Genome scanning of many markers will help to identify individuals with greater amounts of admixture so that they can be removed from the breeding pool</td>
</tr>
<tr>
<td>Identification of units of conservation: species, evolutionarily significant units and management units</td>
<td>The incorporation of adaptive genes and gene expression will augment our understanding of conservation units based on neutral genes. The use of individual-based landscape genetics will help to identify boundaries between conservation units more precisely</td>
</tr>
<tr>
<td>Minimizing adaptation to captivity</td>
<td>Numerous markers throughout the genome could be monitored to detect whether populations are becoming adapted to captivity</td>
</tr>
<tr>
<td>Predicting harmful effects of inbreeding depression</td>
<td>Understanding the genetic basis of inbreeding depression will facilitate the prediction of the effectiveness of purging. Genotyping of individuals at loci associated with inbreeding depression will allow the selection of individuals as founders or mates in captive populations. Pedigree reconstruction will allow more powerful tests of inbreeding depression</td>
</tr>
<tr>
<td>Predicting the intensity of outbreeding depression</td>
<td>Understanding the divergence of populations at adaptive genes will help to predict effects on fitness when these genes are combined. Detecting chromosomal rearrangements will help to predict outbreeding depression</td>
</tr>
<tr>
<td>Predicting the viability of local populations</td>
<td>Incorporating genotypes that affect vital rates and the genetic architecture of inbreeding depression will improve population viability models</td>
</tr>
<tr>
<td>Predicting the ability of populations to adapt to climate change and other anthropogenic challenges</td>
<td>Understanding adaptive genetic variation will help to predict the response to a rapidly changing environment or to harvesting by humans and allow the selection of individuals for assisted migration</td>
</tr>
</tbody>
</table>

From Allendorf et al. (2010)
Conservation genomics gap
Stimulated a discussion

```
Letter
Genomics in Conservation: Case Studies and Bridging the Gap between Data and Application

However, we challenge Shafer et al.'s [1] relatively pessimistic assertion that 'conservation genomics is far from seeing regular application'. Here we illustrate by examples that conservation practitioners utilize more genomic research than is often apparent. In addition, we highlight the work of nonacademic laboratories (government and nongovernmental organizations (NGOs)), some of which are not always well represented in peer-reviewed literature. Finally, we suggest that

2.11. Techno-fix or Tech-no-fix?
Rather than removing the agent of decline, a central concept in conservation, there appears to be a growing culture relying on technological developments to engineer solutions to complex conservation problems, searching for a “silver bullet” solution. Increasing reliance on technological solutions risks becoming a crutch for conservation biology, something ECRs should approach with circumspection. Technology may contribute to solving conservation problems -- and indeed many technologies have great potential and contribute vitally important knowledge to conservation science -- but examples where technology alone has provided true solutions to conservation issues are somewhat limited.

“The conservation genomics gap
Posted on 23 February, 2015 by Bob Denton

MIND THE GAP

"Strong words hopefully beget strong discussion and solutions for this problem."
```
Real challenges remain

• (High) degree of uncertainty remains

• Qualitatively novel aspects of genomics still far from application - *Increased resolution due to marker number is a small component*

• Practitioners still see little use for genetics let alone genomics AND ConGRESS survey at odds with academics with only 1% viewing local adaptation as an important topic
But there are reasons to be optimistic

1. Genomics of natural populations still in its infancy
   - this is an exciting time for basic research

2. New methods and standards are emerging regularly
   - Pipelines, outlier scans, polygenic traits

3. *Real* outliers are often detected regularly
   - But be cautious of story-telling
4. RADseq rapidly replacing microsatellites
   - Standard conservation genetic queries

5. Gene-editing technologies
   - Invasive species and genetic disorders

6. Conservation genomics is still young!

But there are reasons to be optimistic
Thanks!

email: aaronshafer@trentu.ca
Detecting local adaptation is not straightforward

Outlier scans

Bierne et al.'s (2013) perspective highlights risk of false positives

Outlier of anonymous marker(s) - what do you do with it then?

Conservation units
Compare mismatch distances to expectation (0 mismatches)