

ORIGINAL ARTICLE

Genomic toolboxes for conservation biologists

Francesco Angeloni, Niels Wagemaker, Philippine Vergeer and Joop Ouborg

Institute for Water and Wetland Research (IWWR), Department of Molecular Ecology, Radboud University Nijmegen, AJ Nijmegen, The Netherlands

Keywords

conservation genetics, genomics/proteomics, inbreeding.

Correspondence

F. Angeloni, Institute for Water and Wetland Research (IWWR), Department of Molecular Ecology, Radboud University Nijmegen, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands.

Tel.: +31-24-3652919;

fax: +31-24-3652409;

e-mail: f.angeloni@science.ru.nl

Received: 12 October 2011

Accepted: 18 October 2011

First published online: 17 November 2011

doi:10.1111/j.1752-4571.2011.00217.x

Abstract

Conservation genetics is expanding its research horizon with a genomic approach, by incorporating the modern techniques of next-generation sequencing (NGS). Application of NGS overcomes many limitations of conservation genetics. First, NGS allows for genome-wide screening of markers, which may lead to a more representative estimation of genetic variation within and between populations. Second, NGS allows for distinction between neutral and non-neutral markers. By screening populations on thousands of single nucleotide polymorphism markers, signals of selection can be found for some markers. Variation in these markers will give insight into functional rather than neutral genetic variation. Third, NGS facilitates the study of gene expression. Conservation genomics will increase our insight in how the environment and genes interact to affect phenotype and fitness. In addition, the NGS approach opens a way to study processes such as inbreeding depression and local adaptation mechanistically. Conservation genetics programs are directed to a fundamental understanding of the processes involved in conservation genetics and should preferably be started in species for which large databases on ecology, demography and genetics are available. Here, we describe and illustrate the connection between the application of NGS technologies and the research questions in conservation. The perspectives of conservation genomics programs are also discussed.

Introduction

Conservation genetics is an applied science, devoted to the study of genetic and evolutionary patterns and processes within the context of biodiversity conservation (Avice 2010; Frankham 2010). On the one hand, conservation genetics investigates the impact of habitat fragmentation and habitat loss on the genetic structure and its consequences for individual fitness, within and between populations of threatened species. On the other hand, conservation genetics implements genetic tools and approaches to estimate parameters that are important for the design and evaluation of conservation plans, such as past population growth rates, effective population sizes, migration rates, and other demographic parameters.

Conservation genetics strongly focuses on the extinction risks of small and isolated populations, because of the potentially strong impact of genetic drift and inbreeding.

These processes lead to the loss of alleles and genotypic diversity and an increase in homozygosity. Increased homozygosity may subsequently lead to inbreeding depression, a phenomenon defined as a reduction in fitness in inbred individuals as compared to their outbred relatives (Charlesworth and Willis 2009). Reduced genetic diversity within a population may corrupt the potential of the population to evolutionarily adapt to changing environments. Thus, both immediate (i.e., inbreeding depression) and future (i.e., adaptive potential) fitness may be lowered, thereby increasing the probability of extinction of the population, which explains the strong focus of conservation genetics on these processes.

Over the years, conservation genetics has greatly improved our understanding of several conservation issues, including the potential impact of habitat fragmentation on species extinction, the importance of genetics for biodiversity conservation, the genetic requirements for

the management of captive populations and invasive species, and the genetic details of taxonomic distinction (Frankham 2010). Despite the wide variety of topics, in many conservation genetic studies, neutral molecular markers, like amplified fragment length polymorphisms (AFLPs) and microsatellites, are applied. These markers have been widely used to study demography, drift, gene flow, and effective population size (Excoffier and Heckel 2006). Application of these markers in studies of populations of different size of a whole suite of species has provided ample evidence that small populations have on average lower genetic diversity and higher inbreeding levels than large populations (reviewed by Leimu et al. 2006).

Despite its many achievements, this approach also has its limitations. First, conservation genetic studies use only a few markers, in most cases in the order of 5–20 microsatellites or 100–500 AFLPs. Therefore, variation in the order of a few thousand nucleotides is screened, representing the coverage of ca. 0.000001% of the average genome. It is therefore justified to ask how representative the results of a marker study are for the patterns at the level of an entire genome.

Second, microsatellites and AFLPs are assumed to be selectively neutral, even though exceptions are found regularly. This makes them ideal for the study of neutral processes like drift and gene flow and for the study of dispersal, connectivity, and habitat fragmentation effects on local populations. Whether neutral markers are as useful for studying processes like local adaptation, loss of fitness by inbreeding, or potential to adapt to changing environments is, however, subject to debate (e.g., Allendorf et al. 2010; Ouborg et al. 2010a). The question is whether variation in neutral markers is representative for variation in functional genes. As neutral markers (i.e., unlinked to genes) are not affected by selection, their variation may be deviating from the variation in genes that are subjected to selection and are important for local adaptation.

Third, neutral markers are unsuited to settle an ongoing debate in conservation: are species, or populations of species, mainly threatened by environmental change or by habitat fragmentation and its associated genetic consequences? How do environment and genes interact and affect fitness and extinction probabilities? These questions must be addressed by studying the activity of genes rather than (the extent of) sequence variation.

It has recently been argued that the integration of genomic techniques in conservation genetics (i.e., conservation genomics) will make it possible to overcome these limitations (Primmer 2009; Avise 2010; Allendorf et al. 2010; Frankham 2010; Ouborg et al. 2010a,b). The great potential of conservation genomics is for a large part the consequence of the application of next-generation

sequencing (NGS) techniques, at a population scale. Incorporation of NGS into conservation genomics brings two essential novelties as compared to conservation genetics. First, NGS makes it possible to study the genetic variation within and between populations at a level that is more representative for the entire genome, while using tens of thousands (instead of tens or hundreds) of markers. Second, NGS allows for the study of gene activity (rather than sequence alone) as function of population and habitat characteristics.

In the remainder of this paper, we first describe how NGS can help to overcome the three main limitations of the conservation genetic approach. In doing this, we focus on the great potential but also discuss the drawbacks of NGS. Next, we present questions that are raised by the incorporation of genomic techniques in conservation genetics and give examples of how these questions are addressed with NGS. Finally, we discuss the perspectives for the field of conservation genomics.

The genomic toolbox: genome-wide screening

Next-generation sequencing refers to a whole suite of different techniques for rapidly sequencing huge numbers of base pairs. NGS technologies include Roche/454 pyrosequencing, Illumina (formerly Solexa) sequencing, the SOLiD system of ABI, the Polonator G.007, Helicos Heliscope, and the system of Pacific Biosystems (reviewed by Metzker 2010; Zhang et al. 2011). Although these techniques differ in several technical details, they share the feature of randomly sequencing the template DNA, RNA, or cDNA. This results in a huge amount of sequences ('reads') that are assembled to larger units, using bioinformatic algorithms. The technical pros and cons of the various techniques have been thoroughly reviewed elsewhere (Metzker 2010; Zhang et al. 2011). Here, we want to outline how NGS might help in getting around the limitations of using a small set of neutral markers.

Genome-wide coverage

It has been observed that the genome is organized into so-called haplotypes or linkage disequilibrium blocks, within which sequences, including genes, are in linkage disequilibrium. Linkage disequilibrium is defined as the association between alleles that occur together more often than can be accounted for by chance. Within the blocks, recombination is very low or absent (Goldstein and Weale 2001; Wall and Pritchard 2003; Laurie et al. 2007; McVean 2007). In natural populations, the average length of these blocks can vary between 1 and 100 kb (Hohenlohe et al. 2010). Thus, an average genome would contain 10^5 – 10^7 of these blocks. Tens of thousands of markers are

needed to adequately estimate the variation across these blocks. By using NGS, one can identify these markers and screen them for variation within and between populations. Several approaches are available.

One run of whole-genome sequencing (WGS), using NGS, of a single individual will result in $0.5\text{--}15 \times 10^9$ bases, depending on the technology used (Metzker 2010). These bases are part of random reads that will be assembled to longer contiguous sequences, or contigs. Software like MSatFinder (Thurston and Field 2006), SciRoKo (Kofler et al. 2007), msatcommander (Faircloth 2008), and iQDD (Megléczy et al. 2010) allow for rapid screening of all contigs and will typically result in the discovery of several thousand microsatellite loci, including their flanking regions, per NGS run (Tsui et al. 2009; Martin et al. 2010).

Second, when several individuals are sequenced simultaneously, variation among reads at a base pair position indicates the presence of a single nucleotide polymorphism (SNP) in a contig. In a similar way, SNPs can be discovered when sequencing a single individual, at those positions where the individual is heterozygous. Softwares such as GigaBayes (Hillier et al. 2008) and VarScan (Koboldt et al. 2009) and assembly software such as CLC-bio (<http://www.clcbio.com>) are instrumental in detecting SNPs in the NGS data. A typical NGS run will result in the discovery of 50 000+ SNPs (it is advisable to confirm these SNPs by independent resequencing, to reduce the risk of false positives). This discovery rate depends on three parameters: the genetic variation within the sample, the genetic variation within the population, and the coverage depth of the NGS run (i.e., the average number of reads per base pair). The sequencing strategy to be followed depends on the research question. If the aim is to discover a maximum amount of variable SNP markers, simultaneous sequencing of at least 30 individuals of large, variable populations, or individuals of different populations, should be performed (e.g., O'Neil et al. 2010; Angeloni et al. 2011). If, on the other hand, the purpose is to describe (fine-scale) spatial variation, or variation as function of environmental or population differences, samples of the appropriate spatial scale or relevant environments or populations should be sequenced. The amount of variation discovered will likely be lower than in the first strategy but will be more relevant for the question asked.

Regardless of the actual purpose, the representation of SNPs in the data can be enhanced by increasing the coverage depth of the NGS run. To do so, two main strategies can be followed. Both strategies aim at selecting only informative parts out of the entire genome and sequence these exclusively. The first strategy is to sequence the transcriptome (i.e., the expressed gene products), rather

than the genome, of the individuals. The percentage coding DNA in the genome is low, around 2–5% in humans for example. Furthermore, not all genes are expressed in every tissue and at every life stage. The transcriptome of a particular tissue in a particular life stage will therefore be 100–1000 times smaller than the genome itself. Sequencing the transcriptome therefore results in many more reads per contig, which increases the coverage depth and thereby the power to discover SNPs. Every SNP that is discovered is linked to a functional gene.

The second strategy is to reduce the genome by manipulating it prior to sequencing. A promising way to do so is the so-called RAD-tag sequencing procedure (restriction-site-associated DNA tags; Hohenlohe et al. 2010; Fig. 1). RAD-tags are short fragments of DNA, flanking a particular restriction site (Miller et al. 2007). The total DNA is restricted, and adaptors, containing the sequence primers for the specific NGS technique, are ligated to the restriction ends. Next, all DNA is fragmented, and only the fragments containing the adaptors are NGS sequenced. This reduces the genome to be sequenced to only the short flanking regions of restriction sites, of which thousands are present throughout the genome. The coverage depth is increased considerably, and SNPs can be readily detected. Advantages of the RAD-tag procedure are that no reference genome is needed and that detection of SNPs and screening their variation is performed in the same run.

Screening SNP variation

Screening SNP variation can be performed in several ways, including the use of RAD-tags. For example, Hohenlohe et al. (2010) designed the adaptors to restriction ends in such a way that they contained both the sequence primer for the Illumina platform and a sample-specific barcode. Variation at the same SNP position between barcodes allows for the estimation of allele and genotype frequencies. Variation among reads, containing the same barcode, at a particular SNP position indicates that that particular individual is heterozygous for the SNP. The number of samples that can be screened simultaneously depends on the number of barcodes available, as long as each individual will be represented several times in the reads for a particular SNP. If too many individuals are sequenced simultaneously, the coverage depth per individual is reduced to a level that is too low for reliable allele and genotype calling.

There are many other methods to perform SNP genotyping. These methods differ in the technology they are based on, e.g., mass spectrometry (iPLEX; Sequenom, San Diego, CA, USA), quantitative PCR (KASP; KBioscience, Hoddesdon, UK), and hybridization (SNP-chips; several

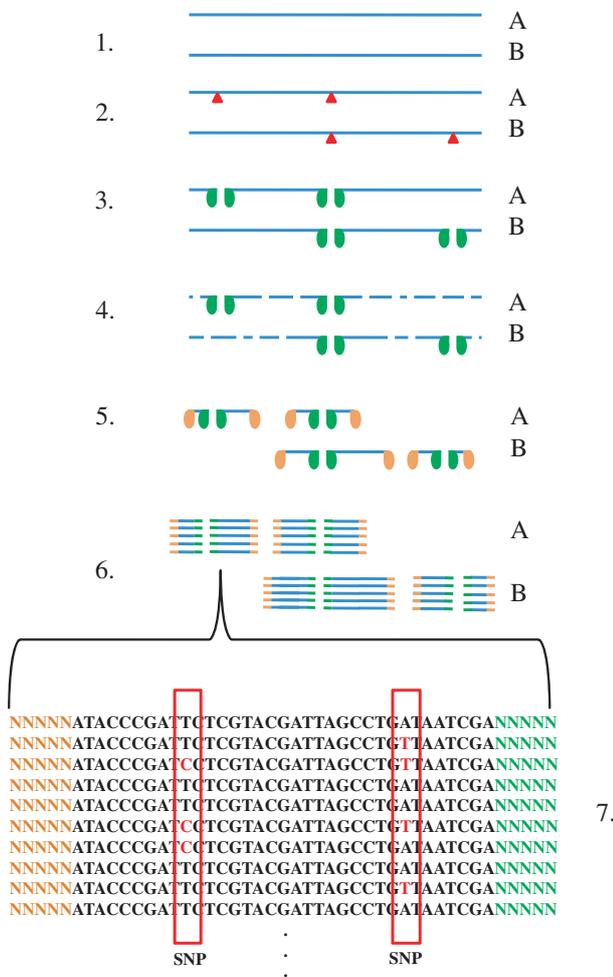


Figure 1 Schematic representation of RAD-tag. In this example, the letters A and B represent two different samples. The first step consists of the isolation of genomic DNA (gDNA) from the samples of interest (i). Then, the gDNA is cleaved with various restriction enzymes (ii). Some of the restriction sites of the enzymes are only present in one sample. A specific adapter (green circles) is ligated to the restriction site (iii), and then, the fragments are sheared (iv). A second adapter is ligated to the other end (v), and a selective PCR step is performed to amplify only the fragments that have adapters on both sides (vi). The adaptors usually contain the Illumina primers and, if necessary, a barcode, so that an Illumina library of barcoded samples can be created. The last step is to perform the sequencing, align reads, and search for variation in SNPs (vii).

manufacturers). These applications also differ in the number of SNP loci that can be simultaneously genotyped: from one or several (KASP, iPLEX) to several hundreds of thousands (SNP-chips). Finally, they differ in the maximum level of multiplexing (i.e., the number of samples that can be screened simultaneously): from around 10–24 (SNP-chips) to more than 1500 (KASP). The choice of a method is based on the tradeoff between investigating a

large number of loci in few individuals or few loci in many individuals. As an example, the Illumina Infinium HD beadchip assays allow for the simultaneous screening of 24 (or 12) samples and 68,000 (or >300 000) SNPs in one run. This technology is based on designing 50-mer probes that represent sequences directly flanking the SNP to be screened and are bound to beads on the chip. After fragmentation, the sample DNA is hybridized to these probes. Next, a one-step enzymatic extension of the probe is performed, in which a labeled nucleotide, that is complementary to the base at the SNP position in the sample, is added. The design of these assays can be customized to any species for which sequences of flanking regions of SNPs are available.

It should be noted that the developments in this field are very rapid, and new techniques, with higher throughput capacity, are becoming available regularly.

The genomic toolbox: functional variation

Processes like genetic drift and gene flow, which are key processes in the conservation of small and isolated populations, will affect every part of the genome in the same way. In contrast, adaptive processes like selection will be gene-specific: only the genes involved in the adaptive responses (and the genes linked to these genes) will be affected and the other (unlinked) genes will respond in a neutral way. To get insight in the potential of a population to adapt to changing environments, we therefore will have to: (i) identify genes and their function and (ii) search for signals of selection in the sequence of these genes, or of markers linked to these genes.

Gene identification

Finding genes becomes relatively straightforward once the reads have been assembled into contigs. The set of contigs can be compared to known sequences in existing databases at NCBI (<http://www.ncbi.nlm.nih.gov/>), by performing a Basic Local Alignment Search Tool search (BLAST). Unfortunately, not all genes will be identified, because the nonmodel species of interest may contain unknown genes, or the sequence of particular genes may have evolved so far that the sequence resemblance with homologues genes in other organisms is too low for BLAST to give a match. Additionally, contigs may be too short to provide sufficient power for matching with known genes. Nevertheless, the procedure provides important knowledge about genes in species that lack other sources of molecular information, as is the case in most species of conservation interest. The resulting database of genes and their functions will serve as reference for studies on the interaction between genes and the environment, for studies on the mechanisms

by which fitness is affected by habitat degradation and habitat fragmentation, and for general functional evaluation of selection signals discovered with the methods outlined below.

Finding signals of selection

One of the main goals of conservation genomics is to identify markers (and genes) under selection and screen populations for variation in these markers (and genes). This aim is shared with evolutionary biology and molecular ecology, where several methods have been described and applied to achieve this goal. Among them are genome-wide selection scans (GWSS; e.g., Storz et al. 2005), genome-wide association studies (GWAS; e.g., Hirschhorn and Daly 2005), and gene–environment association studies (GEA; e.g., Bierne et al. 2011). Although these methods differ in their details, they all rely on screening and interpreting the variation in thousands of SNPs, distributed across the genome.

Selection is a gene-specific process, affecting only specific genes (and any sequence linked to them). For instance, directional selection will reduce the variation in these genes. In contrast, genetic drift is a genome-wide process, affecting all genes and all noncoding DNA in the same way. GWSS therefore compares the variation among SNPs throughout the genome. When a single population is studied, the genetic variation found is averaged over all SNP loci, to create a null distribution of variation. Because the vast majority of markers will behave neutrally, this distribution can be considered to be a null model for selection. Any SNP that shows significantly lower variation than the null model displays a putative signal of selection. When two (or more) populations are compared, a measure of genetic divergence between populations per SNP locus (e.g., F_{ST} , Wright 1984) is averaged over all loci, to create a null distribution. Any SNP whose value significantly deviates from this distribution displays a putative signal of selection.

Genome-wide association studies is an approach based on searching for associations between the variation in SNPs and the variation in phenotype (e.g., Tian et al. 2011). The procedure will result in a set of markers that are associated with a particular phenotypic trait. GEA extends this approach by specifically incorporating the effect of environment on gene–trait associations. It is outside the scope of this paper to discuss the detailed models, statistics, and approaches behind these population genomic procedures; we refer to excellent reviews on this topic (Luikart et al. 2003; Leinonen et al. 2008; Whitlock and Guillaume 2009; Oleksyk et al. 2010).

Once markers with a selection signal have been identified, one of two steps (or both) follows. First, markers

can be used to reconstruct demographic parameters, as it has been done frequently with patterns of microsatellite or AFLP variation (e.g., Ouborg et al. 1999; Bos and DeWoody 2005, Barluenga et al. 2011). The population genetic models to convert an observed marker pattern into an estimate of a demographic parameter are, however, frequently based on the assumption that the markers used are not subjected to selection. Because we now can separate putative neutral from non-neutral markers, the markers with a selection signal can be removed from the data set. The resulting data set, containing only neutral markers, can then be used to reconstruct demographic parameters, like population growth rate, effective population size, and migration rate, using the appropriate population genetic models. The resulting estimates will be unbiased by selection (Luikart et al. 2003). Second, in case we are interested in selective and adaptive processes, we retain only the markers with a selection signal. If these markers are found within a gene, as discovered with the BLAST procedure, this gene is a candidate for being of adaptive importance. In most cases, the marker itself is not under selection but is linked to a gene that is. In that case, a process to find the gene associated with the marker, and responsible for the selection signal, can be started. This is, however, a tedious and risky process, with no guarantee of success (Slate et al. 2009). Luckily, in many cases, it may not be necessary to perform this step. If a marker has been identified as being under selection, screening its variation within and between populations will give insight in functionally important variation, even if the functional details are lacking.

The genomic toolbox: gene activity

Through the use of markers, conservation genetics exclusively focuses on the relationship between landscape and population characteristics, and sequence variation among individuals and populations. Incorporating genomic techniques now makes it possible to investigate the relationship between landscape characteristics and population characteristics, and gene activity variation. The study of gene expression in a conservation context is relevant for the discussion on the importance for population viability of genetics as compared to environment, because gene expression is affected by both genetic and environmental factors. Conservation genomics allows for an approach where genes and environment are treated as interacting entities rather than as exclusive alternative causes of lowered fitness.

Quantification of gene expression can be performed in a number of ways that differ in their suitability for conservation-related studies. For instance, gene expression can be quantified using real-time quantitative PCR

(RT-qPCR). The exact details of this procedure can be found in Derveaux et al. (2010). Unfortunately, for large-scale population studies, this method offers insufficient throughput capacity, but it is a very useful and reliable technique, for instance, in studies on a limited number of candidate genes for adaptation (e.g., Pavli et al. 2011; Prins et al. 2011).

Much higher throughput capacity was achieved with the introduction of microarrays. Microarrays contain tens of thousands of probes that are complementary to expressed genes of a particular species. By hybridizing transcriptome samples (after having reverted the mRNA into cDNA) to a microarray, the expression levels of thousands of genes can be studied simultaneously. Microarray analysis resulted in great progress in gene expression analysis, in the study of genetic linkage, and in association studies (e.g., Kammenga et al. 2007; Zhang et al. 2011). Incorporating microarray technology in the study of nonmodel organisms however offers a major challenge. Within the context of conservation genomics, a major drawback is that, to optimally design the microarray, detailed sequence information of the focal species is needed. In some cases, microarrays of phylogenetically close model species can be used, making the nonmodel species ‘genomically enabled’ (Kohn et al. 2006). In all other cases, genome-wide sequences can be delivered using NGS techniques (i.e., a fully sequenced genome or transcriptome). However, in most cases, this will require more time and money than would be feasible within most conservation programs. More importantly, the NGS data needed to design the microarray may already contain the necessary data to quantify gene expression, as we will outline below. Therefore, when working with a species that is phylogenetically not closely related to a model species, gene expression studies can best be performed with NGS technology.

Despite their limitations, microarrays have been successfully applied in conservation biology. For example, Tymchuk et al. (2010) have estimated the variation in the transcriptome of the threatened Atlantic salmon (*Salmo salar*) to assess the population structure and the potential for adaptive variation. Microarrays have also been applied to study the molecular basis of inbreeding depression. Kristensen et al. (2002) and Pedersen et al. (2005) found increased expression of heat-shock protein 70 (Hsp70) in inbred individuals of *Drosophila melanogaster* and *D. buzzatii*.

NGS-based methods

Next-generation sequencing can be used to study differences in gene expression patterns, in a procedure referred to as RNA-seq (Wang et al. 2008; Wilhelm and

Landry 2009; Marguerat and Bähler 2010). This approach is based on NGS of the transcriptome (Fig. 2). The resulting reads are assembled to contigs, to create or extend a reference transcriptome. Next, the individual reads are mapped on this reference transcriptome. The obtained number of reads per contig is a measure of the expression level of the corresponding gene. If different transcriptome samples can be labeled with a barcode, in the preceding PCR process necessary to create the library to be sequenced, many samples can be sequenced and mapped simultaneously in the same run. Each next run represents a new sample of the same transcriptome. Therefore, each new run is first assembled to the exist-

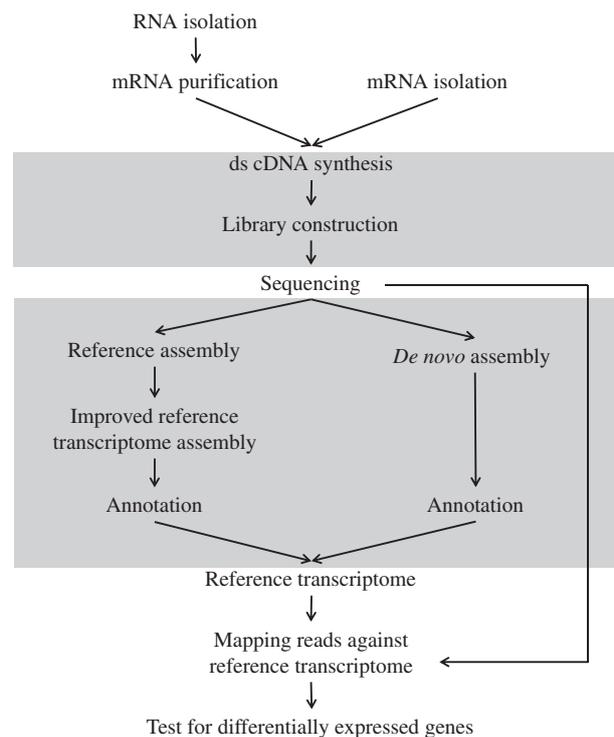


Figure 2 Pipeline representation of an RNA-seq experiment. Grey areas represent optional steps. After mRNA isolation or purification from several individuals or treatments, double-strand cDNA is synthesized, resulting in an EST library. This library is then sequenced using the desired next-generation sequencing technology. Next-generation sequencing produces a large amount of reads, which are usually assembled using a reference genome or transcriptome. The new sequencing information can be added to the reference to improve its quality. If a reference transcriptome is not available, reads are assembled *de novo*. In this case, the *de novo* contigs are usually annotated, to retrieve functional information. Once a reference transcriptome is available, reads produced by sequencing are mapped against the reference transcriptome, and the number of reads per contig per sample is expressed in a standardized way (e.g., Reads Per Kilobase of contig per Million mapped reads, see text). The last step is to test whether difference in expression values between samples is significant.

ing reference transcriptome, before being mapped on the new reference transcriptome. As reads of 50–100 bp will be reliably mapped to the corresponding contig, for RNA-seq obtaining many short reads is more helpful than obtaining fewer longer reads. Therefore, until now, RNA-seq is almost exclusively performed on the Illumina platform. Metzker (2010) pointed out that PCR bias, which may result from the procedure to create the Illumina library, may lead to biased expression estimates. Therefore, he advised to perform future RNA-seq on systems that do not rely on PCR, such as the Helicos system. This system can even sequence RNA directly rather than first converting mRNA into cDNA (Metzker 2010).

The results of an RNA-seq experiment will be counts of reads per contig. Usually, this value is expressed as a standardized value called Reads Per Kilobase of contig per Million mapped reads (RPKM; Mortazavi et al. 2008). In experiments aimed to compare the number of reads between samples, proportion tests are applicable (Auer and Doerge 2010). This type of test compares the gene expression levels as proportions rather than raw counts. Therefore, the data are corrected for sample size. Several types of proportion-based tests exist: Fisher's exact test (Fisher 1935), Kal et al.'s test (1999), and Audic and Claverie's test (1997). Comparison of these tests showed only marginal performance differences (Man et al. 2000; Romualdi et al. 2001; Ruijter et al. 2002).

NGS: challenges

While NGS offers many exciting opportunities for conservation genomics, there are a number of challenges as well. A serious challenge is the high running cost of an NGS study. Even though the price per bp sequence has decreased considerably over the last few years (de Magalhães et al. 2009), the costs are still too high to perform large-scale population studies. Developments are going fast, and it is expected that within a few years the prices will go down to a level that makes NGS affordable for many conservation programs. Nevertheless, the high costs currently force researchers to make a choice between two extreme ends of the research spectrum: spending the available money on screening many individuals with a few (e.g., twenty) microsatellites or on screening a few individuals with NGS. In the first case, much information at the population and metapopulation level will be obtainable, but the information per individual will be quite limited. In the second case, very detailed information per individual will become available, which will however be difficult to extent to higher-than-individual organizational levels. The choice should fully depend on the research question asked.

A second point is that performing NGS is technically quite demanding. While the questions often will be formulated within the community of conservationists and ecologists, the translation of these questions to actual NGS experiments requires close cooperation with genomicists and bioinformaticians. Optimal communication between applied conservationists, conservation geneticists, and conservation genomicists is required to fully exploit the potential of NGS in a conservation context. Attempts are currently being made to create user interfaces to facilitate communication between these groups (e.g., the CONGRESS project <http://www.congressgenetics.eu>).

A third issue is that NGS produces huge amounts of data. Each run of a given NGS platform can sequence up to hundreds of Gbp of DNA (Metzker 2010). An average NGS experiment, be it WGS, transcriptome sequencing, SNP analysis, or an RNA-seq experiment, will easily produce hundreds of gigabytes of raw data. While this is one of the attractive features of NGS, at the same time, it poses a serious challenge for data management. Application of NGS at least involves setting up a massive data storage facility and a bioinformatics pipeline to effectively analyze the sequence data. Fortunately, there exists a wide variety of software tools capable of processing NGS data. An extensive list of bioinformatic tools, with links to the respective websites, can be found in Zhang et al. (2011). Performing population genetic analyses with thousands of markers is also not immediately straightforward. Therefore, close cooperation with bioinformaticians, or with laboratories that have extensive expertise to perform both NGS and the corresponding bioinformatic analyses, is strongly advisable.

Application of NGS in a conservation context

This review started with identifying three main issues in conservation genetics that could be solved by taking a conservation genomics approach: genome-wide screening of genetic variation, distinguishing between neutral and non-neutral variation, and assessing variation in gene activity rather than in gene sequence alone. The application of (various forms of) NGS now allows for addressing these issues in full detail. Two important questions remain to be answered: why and when is taking up a conservation genomics approach, and using NGS, most profitable? And how to proceed and which techniques must be used, when the aim is to answer specific questions? We address both questions subsequently.

Why and when to use a conservation genomic approach?

On the one hand, conservation genetics is a discipline that performs fundamental research on how population genetic

processes determine and interact with demographic processes. For instance, what is the dynamics of genetic variation in populations of varying size, and how can we explain these dynamics? On the other hand, conservation genetics is a discipline that performs applied research on conservation issues. For instance: is there evidence that the population we want to conserve suffers from loss of genetic variation or inbreeding? Or how much dispersal is there between the populations in a particular landscape, and are infrastructural changes, like the establishment of corridors, efficient in enhancing dispersal? These two lines within conservation genetics mutually influence each other, nevertheless their goals are different.

The priority aim of taking up a conservation genomic approach is to better understand the relevant conservation genetic processes. Therefore, conservation genomics is expanding the fundamental research side of conservation genetics. For the applied side, it seems less urgent to use conservation genomics. Many of the applied questions are currently addressed in a satisfactory way using microsatellites or AFLPs. This may however change, if conservation genomic studies demonstrate that conclusions based on the application of thousands of SNPs qualitatively and quantitatively differ from conclusions based on a small number of microsatellites. Nevertheless, taking up conservation genomics and NGS is at the moment most acute in the fundamental research line of conservation genetics.

The application of NGS will result in large amounts of data on the details of genomes. These data will only become most valuable if they can be interpreted in the context of conservation and general ecology. For such an interpretation, extensive data sets on ecology and demography of the populations under study are needed. Therefore, as it has been pointed out previously (Primmer 2009; Ouborg et al. 2010a), a conservation genomics program will be most profitable in a species that has been amply investigated on ecological, demographic, and genetic aspects.

Match between question and approach

Even though NGS is a relatively new technology that has only recently been applied in ecology and conservation biology, several studies demonstrate its large potential. In most studies, NGS technology is used to identify genes of importance for conservation. For instance, the California condor has a relatively high frequency of an inheritable dwarfism called chondrodystrophy (Ralls et al. 2000; Ralls and Ballou 2004). NGS technologies are currently being applied to identify carriers of the disease, which offers the opportunity to eliminate the disease (Romanov et al. 2006; Frankham 2010).

Several studies have used NGS technology to characterize the transcriptome of species with conservation interest. The Glanville fritillary butterfly (*Melitaea cinxia*) was one of the first nonmodel species for which a large part of the transcriptome was characterized, using a Roche/454 platform (Vera et al. 2008). The authors characterized around 9000 unique genes, with an average coverage of 6.5-fold for the 4800 longest contigs. This coverage was sufficient for the identification of a large number of SNPs, including 149 first and second codon position polymorphisms, which are likely to change the corresponding amino acid sequence. The genomic resources described in Vera et al. (2008) enable the study of ecological features of *M. cinxia* (e.g., dispersal ability). In a follow-up study, Wheat et al. (2011) combined the genomic resources developed by Vera and colleagues with a long-term ecological study to obtain a more mechanistic understanding of life history variation affecting ecological and evolutionary dynamics of *M. cinxia*. The authors first identified groups of populations that differed in their demographic history. Gene expression differences and allelic polymorphisms were subsequently linked with life history traits and population dynamics to identify new candidate genes that affect eco-evolutionary dynamics. Their results are important for the conservation of *M. cinxia*, as the life history traits they studied are known to affect metapopulation persistence in fragmented habitats (Hanski and Ovaskainen 2000).

Soon after the Vera et al. study, several other studies that characterized the transcriptome of several nonmodel species followed (e.g., Novaes et al. 2008; Kristiansson et al. 2009; Meyer et al. 2009; Parchman et al. 2010; Angeloni et al. 2011).

There are many applications of NGS, and the number of applications is increasing continuously. Starting a conservation genomics research program therefore involves both formulating very precise questions and finding a match between question and NGS application. Here, we discuss this match in three categories of questions that can most profitably be addressed with a conservation genomic approach (Fig. 3).

1 Genome-wide screening of genetic variation.

One important question in conservation genetics is whether patterns of markers are accurate estimations of processes like drift, inbreeding, and gene flow. Applying NGS allows for the investigation into whether patterns of genome-wide variation, as measured with thousands of SNPs, lead to the same conclusions about these population genetic processes as patterns of the variation in 10–20 microsatellites. For instance, the genome-wide estimation of heterozygosity across all SNPs is negatively correlated with the level of individual inbreeding (Keller and Waller 2002), and SNP variation might therefore

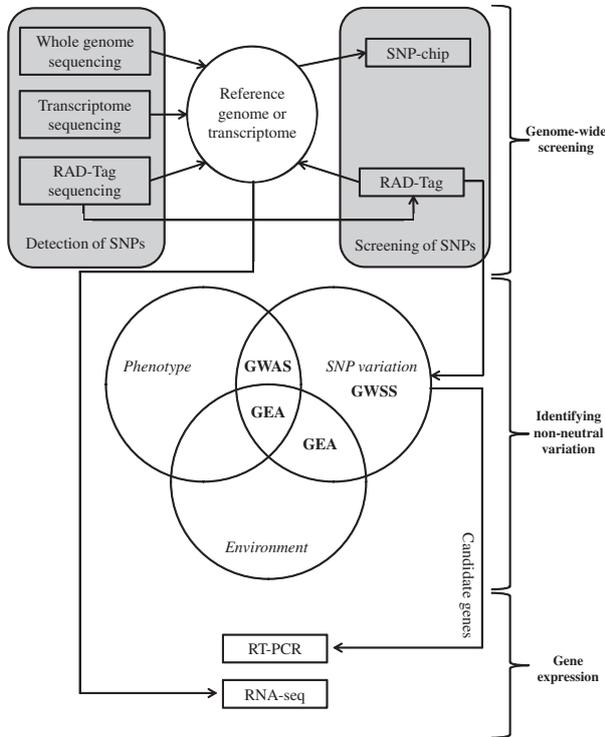


Figure 3 A scheme of how various next-generation sequencing approaches relate to the three main categories of questions in conservation genomics and how to feed their results into each other.

provide a more accurate estimate of inbreeding. Moreover, using large numbers of SNPs, distributed across the genome, is expected to lead to more adequate estimates of population genetic parameters (Novembre and Stephens 2008), to easier detection of signals of selection (Slate et al. 2009), to more power in assigning individuals to parents or other kin (Santure et al. 2010), and to estimates of historical demography (Ekblom and Galindo 2010).

The choice between various approaches to detect and screen SNPs depends on the final goal and the resources available. If the goal is to perform a single experiment to screen a population of a nonmodel species on SNP variation, performing an RAD-tag experiment seems to provide the best balance between level of detail in the data and costs and efforts invested. If the goal is to develop SNPs for many follow-up experiments, it is better to perform WGS or transcriptome sequencing, so as to create a reference genome that can be annotated. Additional advantage of these approaches is also that other types of markers, most notably microsatellites, can be identified in the same run. The developed SNPs can then be screened either with a SNP-chip that is designed based on these results or with an RAD-tag sequencing procedure. In the

latter case, because now a (annotated) reference genome is available, the SNP variation discovered by RAD-tag sequencing can be mapped on the genome and be functionally evaluated.

All described approaches have been applied to non-model species and in ecological or conservation contexts. Sanchez et al. (2009), in an effort to develop genomic tools for the Rainbow trout (*Oncorhynchus mykiss*), performed WGS on a pool of genomic DNA composed out of 96 unrelated rainbow trout. Three independent analyses were performed on the data, resulting in the identification of 22 022–47 128 putative SNPs.

Novaes et al. (2008) used 454 pyrosequencing to characterize the transcriptome of *Eucalyptus grandis*, the most widely planted hardwood tree species. They used RNA of vegetative tissues sampled from 21 different genotypes and detected 23 742 SNPs, 83% of which were then validated after resequencing. This information was then used to detect evolutionary signatures of genes by studying nonsynonymous and synonymous substitutions. Therefore, several genes were discovered that are under purifying selection.

Angeloni et al. (2011) sequenced the transcriptome of 48 individuals of the locally threatened plant species *Scabiosa columbaria*, using a combination of 454 and Illumina sequencing. They found a total of 75 054 putative SNPs. They also identified 4320 microsatellites, for which 856 had suitable flanking regions for primer design.

To study the parallel evolution of marine and freshwater populations of the three-spined stickleback (*Gasterosteus aculeatus*), Hohenlohe et al. (2010) applied RAD-tag sequencing on an Illumina platform to simultaneously detect and genotype SNPs. They identified over 45 000 SNPs in two oceanic and three freshwater natural populations of threespine stickleback, for a total of 100 individuals. Further analyses showed that these SNPs were evenly distributed across the entire genome. Several chromosomal regions in stickleback were found that were highly differentiated between the two ecotypes. These regions contained both previously identified loci of large phenotypic effect and novel candidate genes involved in stickleback phenotypic evolution.

The same technology was applied by Hohenlohe et al. (2011) to identify almost 3000 candidate SNP loci with fixed allelic differences between introduced rainbow trout (*Oncorhynchus mykiss*) and native west-slope cutthroat trout (*Oncorhynchus clarkii lewisi*), using a total of 24 individuals.

Rowe et al. (2011) present a review on the application of RAD-tag sequencing, with NGS, in different fields.

2 Adaptation and genotype X environment.

Another important goal of conservation genomics is to study the interaction between genes and their

environment, in a conservation context. NGS allows for the study of the balance between genetic effects of habitat fragmentation (inbreeding, loss of genetic variation) and effects of habitat degradation, at the genomic level. NGS also allows for the identification of the genes involved in adaptation. Methods following NGS, like GWSS and GWAS, allow for distinguishing neutral from non-neutral markers and thus for screening of the effect of habitat fragmentation on patterns of non-neutral (as compared to neutral) marker variation.

Stapley et al. (2010) provided an excellent overview of NGS approaches in the study of adaptation. In short, the first step is to create a dense map of SNP markers across the genome. This can be done with WGS, with transcriptome sequencing, or with NGS of targeted (candidate) regions. Also RAD-tag sequencing could be used, although this may deliver less dense maps. Based on screening many individuals for thousands of SNPs, one or more of the following approaches can be used to identify the loci involved in adaptation. A GWSS procedure analyzes only SNP data and identifies outlier loci, as candidate areas involved in adaptation. A reference genome is not an absolute requirement, as the method only searches for markers with a deviating level of variation. Therefore, GWSS can be performed in nonmodel species that lack a reference genome, using RAD-tag sequencing. However, only when a reference genome is available can the identified markers be associated with areas in the genome, which is the starting point for further functional analyses. If besides variation in SNPs, also variation in phenotype is assessed, associations between markers and traits can be found in a GWAS procedure. This does, however, require a reference genome, where the position of markers relative to each other is known. In some cases, species of conservation interest can be studied using reference genomes of closely or more distantly related model species, as linkage groups are likely to be conserved across related species. Alternatively, transcriptome analysis using an RNA-seq procedure can be used to identify genes that are associated with differences between populations, be it genetic or environmental differences (or both).

Next-generation sequencing provides many advantages in this type of research (Stapley et al. 2010). For instance, it provides much more power, using more loci and more individuals, thereby facilitating the discovery of selection signals or of loci of small effect. Perhaps, the biggest advantage is that the dynamics of genes involved in adaptation can now be evaluated within the context of the dynamics of other parts of the genome. This opens the way to separating effects of genetic drift from effects of selection, and effects of selection from effects of demography. Eventually, this will allow us to investigate what the balance is between genetic drift and

local adaptation, in small populations or in systems of isolated populations.

Genome-wide selection scans was performed by Galindo et al. (2010), who applied 454 pyrosequencing to characterize the transcriptome of two different ecotypes of the marine gastropod *Littorina saxatilis*. This gastropod is a good species to study ecological speciation. Galindo and colleagues collected 15 females per ecotype in each of the two sampling site. Females were pooled into two samples, each with 30 individuals (one sample per ecotype). Two thousand four hundred and fifty-four SNPs were found, 7% of which were identified as outliers that may represent direct targets of selection or regions tightly linked to selected loci.

Atwell et al. (2010) applied GWAS to study the genetics of 107 phenotypes of *Arabidopsis thaliana*. Several adaptive traits, including flowering time and pathogen resistance, were shown to be controlled by loci of major effect. The study also showed that it may be difficult to distinguish between true association and false positives because of the confounding effect of population structure (see also Bierne et al. 2011). Nevertheless, the authors demonstrated that GWAS can be successfully performed on *Arabidopsis* and can also be applicable in other, non-model organisms.

Turner et al. (2010) performed a GEA study, where they investigated whether *Arabidopsis lyrata* is locally adapted to serpentine soil, by mapping the polymorphisms responsible for such adaptation. They pooled approximately 200 DNA samples extracted from individuals from serpentine and nonserpentine soil and sequenced each pool with Illumina. The polymorphic SNPs that were most strongly associated with soil type were involved with heavy metal detoxification and calcium and magnesium transport. These SNPs provide several candidate polymorphisms for adaptation in serpentine soil. The authors then confirmed the results by sequencing three candidate loci in the European subspecies of *A. lyrata*, finding parallel differentiation of the same polymorphism at one locus.

3 The study of mechanisms.

NGS will be instrumental in the study of the mechanisms underlying the relationship between genetic effects of habitat fragmentation and the final consequences for fitness and population viability. Inbreeding depression, the reduced fitness of offspring from a mating between related individuals, plays a central role in conservation biology. The average level of inbreeding in small and isolated populations is expected to increase over time, making individuals more homozygous, which leads to increased expression of recessive deleterious alleles and reduced fitness. NGS technologies make it possible to study the genetic architecture of inbreeding depression

(Kristensen et al. 2010). One way to proceed here is to screen SNP variation in a large number of individuals that differ in inbreeding level. If for each individual also fitness traits are measured, associations between inbreeding level, SNP markers and fitness traits can be assessed (Kristensen et al. 2010). In another approach, differential expression of genes between inbred and outbred individuals can be investigated in an RNA-seq procedure. This would pinpoint genes that are associated with inbreeding depression, either as cause or as consequence. In controlled environment studies with inbred and outbred individuals, the nature of the interaction between inbreeding depression and environmental stress (Armbruster and Reed 2005) can be elucidated.

Although this type of work is in its infancy, the first results underline the need for a conservation genomic approach (Ayroles et al. 2009). Lippman and Zamir (2007) reviewed results that show that inbreeding depression is generally based on the action of several loci but is not associated with genome-wide heterozygosity in regions outside these loci. In a series of microarray experiments with *Drosophila melanogaster* (Kristensen and Sørensen 2005; Kristensen et al. 2006; Pedersen et al. 2008), it was shown that different populations may have different genetic causes of inbreeding depression. In a series of RNA-seq experiments with the plant species *Scabiosa columbaria*, it was shown that inbreeding depression in different genotypes may be caused by different genes (Angeloni et al., unpublished research). On the other hand, despite these differences in both studies, there was also a general response. In *D. melanogaster*, genes involved in stress responses generally respond to inbreeding (Kristensen et al. 2006). In *S. columbaria*, the first results indicate that genes involved in energy metabolism respond to inbreeding (F. Angeloni, N. Wagemaker, J. Ouborg, unpublished data). Studies on genetic architecture and mechanisms of important conservation genetic processes like inbreeding depression, using NGS approaches, are just starting to emerge, and many exciting and new results are expected in the near future.

Other examples of the application of RNA-seq include studies on birds and fishes. Ekblom et al. (2010) investigated tissue-specific gene expression patterns in the zebra finch (*Taeniopygia guttata*). In particular, they examined genes of the major histocompatibility complex (MHC). MHC genes are among the most thoroughly studied example of adaptive molecular evolution. The authors sequenced and assembled RNA from six different tissues, for a total of 11 793 ESTs. They found evidence for tissue-specific differential expression of 10 different genes related to MHC, primarily in spleen and brain.

Künstner et al. (2010) used RNA-seq for a comparative genomic study of the avian genome. The authors

sequenced the brain transcriptome of 10 different non-model bird species and identified nearly 6500 genes. Among other results, they found evidence for a higher mutation rate of the Z chromosome when compared to autosomes. Overall, their study demonstrates the usefulness of NGS technologies for comparative genomic analysis for nonmodel species.

Elmer et al. (2010) performed RNA-seq to examine transcriptome differences between ecologically divergent, endemic and sympatric species of cichlid fishes (*Amphilophus astorquii* and *Amphilophus zalius*). The authors identified six genes showing signals of strong diversifying selection. These genes were involved in biosynthesis, metabolic processes, and development. NGS technologies enabled the authors to infer that natural selection is acting to diversify the genomes of young species, such as cichlids, to a much larger extent than was previously thought.

Perspectives

In this paper, we have discussed the great potential of conservation genomics and the application of NGS technology. NGS should be able to overcome three major limitations of conservation genetics by providing genome-wide screening, offering insight in functional genetic variation, and integrating environmental and genetic parameters in the analysis of gene expression.

The approaches outlined here are based on the present state of the NGS technology. However, developments are taking place at a breathtaking speed. What is the best choice of platform now may not be the best choice in half a year time. The amount of data produced and the throughput capacity of samples are increasing ever further, while at the same time the costs are decreasing. Although we are not there yet, it is foreseeable that in the years to come it will become feasible to sequence individuals completely, instead of relying on markers. This would further increase the value of NGS for areas like ecology and conservation biology.

Having said that, the most important question would still be: what new insights will the application of these techniques bring us? Obviously, the answer to this question can only be given in due time, but we anticipate major new insights when answering the following questions.

The first set of questions concern the balance between neutral and functional variation. Is what we have measured with neutral markers representative for the variation in functional genes? What is the balance between genetic drift and natural selection, or in other words: at what size is population size (i.e., drift) more important for fitness and population viability than habitat quality

(i.e., selection)? To answer these questions, we can now develop thousands of markers that cover the entire genome, apply them in a population genomic approach, and compare the resulting signals of selection, or the variation in associated genes, among populations of varying size.

The second series of questions concern the insights that might be obtained by analyzing gene expression level rather than gene sequence variation. Is what we have measured and concluded with neutral markers representative for what happens at the gene expression level? How is the heritability of gene expression related to population size (Visscher et al. 2008)? Is gene expression a better predictor for fitness than (neutral or functional) markers? These questions may be answered after broad-scale application of gene expression assays, such as RNA-seq. A whole suite of experiments within this context is waiting and ready to be performed.

The third type of questions concerns the interaction between genetics and environment. Can we estimate the relative contribution of genetics to fitness, in particular in comparison with the contribution of environment? How do habitat fragmentation, genetic drift, and inbreeding affect the interaction between genetics and environment? Gene expression studies in populations of varying size and environmental quality will provide insight here.

The fourth category of questions is whether we can understand the mechanisms of processes like inbreeding depression and thereby anticipate its likely impact rather than measure its impact a posteriori. How many genes are involved in inbreeding depression? Which genes are involved? What is their relative contribution? How is their expression affected by environmental factors? Can we use their sequence, or the allelic state of associated SNPs, as proxy for future inbreeding depression? Here, we need the whole suite of genomic approaches outlined in this paper. We need to have a reference genome, we need to characterize the transcriptome, and we need differential expression studies to identify genes associated with fitness effects of inbreeding. Answering these questions is not a trivial task at all, but if we want to understand how genetics and environment might affect future population viability, we need to face the challenge.

Finally, can we identify units of conservation in a more meaningful and accurate way when the decision is based on assessments of functional rather than neutral variation? Comparing spatial patterns of neutral and functionally associated markers will shed light on this issue.

In all these cases, conservation genomics and the application of NGS will be indispensable. Evolving from conservation genetics to conservation genomics is not merely an extension of existing approaches, but it opens the way to asking and answering totally new questions. Exciting results will be obtained in the coming years, and they will

put the conservation genetics paradigm (Ouborg et al. 2006) to a test.

Literature cited

- Allendorf, F. W., P. A. Hohenlohe, and G. Luikart. 2010. Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11**:697–709.
- Angeloni, F., N. Wagemaker, M. Jetten, H. den Camp, E. Janssen-Megens, K. François, H. Stunnenberg *et al.* 2011. De novo transcriptome characterization and development of genomic tools for *Scabiosa columbaria* L. using next generation sequencing techniques. *Molecular Ecology Resources* **11**:662–674.
- Armbruster, P., and D. H. Reed. 2005. Inbreeding depression in benign and stressful environments. *Heredity* **95**:235–242.
- Atwell, S., Y. Huang, B. J. Vilhjalmsón, G. Willems, M. Horton, Y. Li, D. Meng *et al.* 2010. Genome-wide association study of 107 phenotypes in a common set of *Arabidopsis thaliana* in-bred lines. *Nature* **465**:627–631.
- Audic, S., and J. Claverie. 1997. The significance of digital gene expression profiles. *Genome Research* **7**:986–995.
- Auer, P. L., and R. W. Doerge. 2010. Statistical design and analysis of RNA sequencing data. *Genetics* **185**:405–416.
- Avis, J. C. 2010. Perspective: conservation genetics enters the genomic era. *Conservation Genetics* **11**:665–669.
- Ayroles, J. F., K. A. Hughes, K. C. Rowe, M. M. Reedy, S. L. Rodriguez-Zas, J. M. Drnevich, C. M. Cáceres *et al.* 2009. A genomewide assessment of inbreeding depression: gene number, function, and mode of action. *Conservation Biology* **23**:920–930.
- Barluenga, M., F. Austerlitz, J. A. Elzinga, S. Teixeira, J. Goudet, and G. Bernasconi. 2011. Fine-scale spatial genetic structure and gene dispersal in *Silene latifolia*. *Heredity* **106**:13–24.
- Bierne, N., J. Welch, E. Loire, F. Bonhomme, and P. David. 2011. The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**:2044–2072.
- Bos, D. H., and J. A. DeWoody. 2005. Molecular characterization of major histocompatibility complex class II alleles in wild tiger salamanders (*Ambystoma tigrinum*). *Immunogenetics* **57**:775–781.
- Charlesworth, D., and J. H. Willis. 2009. The genetics of inbreeding depression. *Nature Reviews* **10**:783–796.
- Derveaux, S., J. Vandesompele, and J. Hellemans. 2010. How to do successful gene expression analysis using real-time PCR. *Methods* **50**:227–230.
- Eklblom, R., and J. Galindo. 2010. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**:1–15.
- Eklblom, R., C. N. Balakrishnan, T. Burke, and J. Slate. 2010. Digital gene expression analysis of the zebra finch genome. *BMC Genomics* **11**:219.
- Elmer, K. R., S. Fan, H. M. Gunter, J. C. Jones, S. Boekhoff, S. Kukaku, and A. Meyer. 2010. Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. *Molecular Ecology* **19**(Suppl. 1):197–211.
- Excoffier, L., and G. Heckel. 2006. Computer programs for population genetics data analysis: a survival guide. *Nature Reviews. Genetics* **7**:745–758.
- Faircloth, B. C. 2008. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* **8**:92–94.

- Fisher, R. A. 1935. The logic of inductive inference. *Journal of the Royal Statistical Society* **98**:39–82.
- Frankham, R. 2010. Challenges and opportunities of genetic approaches to biological conservation. *Biological Conservation* **143**:1919–1927.
- Galindo, J., J. W. Grahame, and R. K. Butlin. 2010. An EST-based genome scan using 454 sequencing in the marine snail *Littorina saxatilis*. *Journal of Evolutionary Biology* **23**:2004–2016.
- Goldstein, D. B., and M. E. Weale. 2001. Population genomics: linkage disequilibrium holds the key. *Current Biology* **11**:R576–R579.
- Hanski, I., and O. Ovaskainen. 2000. The metapopulation capacity of a fragmented landscape. *Nature (London)* **404**:755–758.
- Hillier, L. W., G. T. Marth, A. R. Quinlan, D. Dooling, G. Fellw, D. Barnett, P. Fox *et al.* 2008. Whole-genome sequencing and variant discovery in *C. elegans*. *Nature Methods* **5**:183–188.
- Hirschhorn, J. N., and M. J. Daly. 2005. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* **6**:95–108.
- Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W.A. Cresko. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**:1–23.
- Hohenlohe, P. A., S. J. Amish, J. M. Catchen, F. W. Allendorf, and G. Luikart. 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources* **11**:117–122.
- Kal, A. J., A. J. van Zonneveld, V. Benes, M. van den Berg, M. G. Koerkamp, K. Albermann, N. Strack *et al.* 1999. Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources. *Molecular Biology of the Cell* **10**:1859–1872.
- Kammenga, J. E., M. A. Herman, N. J. Ouborg, L. Johnson, and R. Breitling. 2007. Microarray challenges in ecology. *Trends in Ecology and Evolution* **22**:273–279.
- Keller, L., and D. Waller. 2002. Inbreeding effects in wild populations. *Trends in Ecology and Evolution* **17**:230–241.
- Koboldt, D. C., K. Chen, T. Wylie, D. E. Larson, M. D. McLellan, E. R. Mardis, G. M. Weinstock *et al.* 2009. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**:2283–2285.
- Kofler, R., C. Schlötterer, and T. Lelley. 2007. SciRoKo: a new tool for whole genome microsatellite search and investigation. *Bioinformatics* **23**:1683–1685.
- Kohn, M. H., W. J. Murphy, E. A. Ostrander, and R. K. Wayne. 2006. Genomics and conservation genetics. *Trends in Ecology and Evolution* **21**:629–637.
- Kristensen, T. C., and A. C. Sørensen. 2005. Inbreeding: lessons from animal breeding, evolutionary biology, and conservation genetics. *Animal Science* **80**:121–133.
- Kristensen, T. N., J. Dahlgaard, and V. Loeschcke. 2002. Inbreeding affects Hsp70 expression in two species of *Drosophila* even at benign temperatures. *Evolutionary Ecology Research* **4**:1209–1216.
- Kristensen, T. N., P. Sørensen, M. Kruhøffer, K. S. Pedersen, and V. Loeschcke. 2006. Inbreeding by environmental interactions affect gene expression in *Drosophila melanogaster*. *Genetics* **173**:1329–1336.
- Kristensen, N. T., K. S. Pedersen, C. J. Vermeulen, and V. Loeschcke. 2010. Research on inbreeding in the 'omic' era. *Trends in Ecology and Evolution* **25**:44–52.
- Kristiansson, E., N. Asker, L. Forlin, and D. G. J. Larsson. 2009. Characterization of the *Zoarces viviparus* liver transcriptome using massively parallel pyrosequencing. *BMC Genomics* **10**:345.
- Künstner, A., J. B. Wolf, N. Backström, O. Withney, C. N. Balakrishnan, L. Day, S. V. Edwards *et al.* 2010. Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. *Molecular Ecology* **19**(Suppl. 1):266–276.
- Laurie, C. C., D. A. Nickerson, A. D. Anderson, B. S. Weir, R. J. Livingston, M. D. Dean, K. L. Smith *et al.* 2007. Linkage disequilibrium in wild mice. *PLoS Genetics* **3**:e144.
- Leimu, R., P. Mutikainen, J. Koricheva, and M. Fischer. 2006. How general are positive relationship between plant population size, fitness and genetic variation? *Journal of Ecology* **94**:942–952.
- Leinonen, T., R. B. O'hara, J. M. Cano, and J. Merila. 2008. Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *Journal of Evolutionary Biology* **21**:1–17.
- Lippman, Z. B., and D. Zamir. 2007. Heterosis: revisiting the magic. *Trends in Genetics* **23**:60–66.
- Luikart, G., P. R. England, D. Tallmon, S. Jordan, and P. Taberlet. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* **4**:981–994.
- de Magalhães, J. P., C. E. Finch, and G. Janssens. 2009. Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions. *Ageing Research Reviews* **9**:315–323.
- Man, M. X., X. Wang, and Y. Wang. 2000. POWER_SAGE: comparing statistical tests for SAGE experiments. *Bioinformatics* **16**:953–959.
- Marguerat, S., and J. Bähler. 2010. RNA-seq: from technology to biology. *Cellular and Molecular Life Sciences* **97**:569–579.
- Martin, J. F., N. Pech, E. Meglécz, S. Ferreira, C. Costedoat, V. Dubut, T. Malausa *et al.* 2010. Representativeness of microsatellite distributions in genomes, as revealed by 454 GS-FLX Titanium pyrosequencing. *BMC Genomics* **11**:560.
- McVean, G. 2007. The structure of linkage disequilibrium around a selective sweep. *Genetics* **175**:1395–1406.
- Megléc, E., C. Costedoat, V. Dubut, A. Gilles, T. Malausa, N. Pech, and J. Martin. 2010. QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* **26**:403–404.
- Metzker, M. L. 2010. Sequencing technologies – the next generation. *Nature Reviews* **11**:31–46.
- Meyer, E., G. V. Aglyamova, S. Wang, J. Buchanan-Carter, D. Abrego, J. K. Colbourne, B. L. Willis *et al.* 2009. Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFLX. *BMC Genomics* **10**:219.
- Miller, M. R., J. P. Dunham, A. Amores, W. A. Cresko, and E. A. Johnson. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* **17**:240–248.
- Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**:621–628.
- Novaes, E. D., R. Drost, W. G. Farmerie, G. J. Pappas, D. Grattapaglia, R. R. Sederoff, and M. Kirst. 2008. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics* **9**:312.
- Novembre, J., and M. Stephens. 2008. Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics* **40**:646–649.
- Oleksyk, T. K., M. W. Smith, and S. J. O'Brien. 2010. Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society B* **365**:185–205.

- O'Neil, S. T., J. D. K. Dzurisin, R. D. Carmichael, N. F. Lobo, S. J. Emrich, and J. J. Hellman. 2010. Population-level transcriptome sequencing of nonmodel organisms *Erynnis propertius* and *Papilio zelicaon*. *BMC Genomics* **11**:310.
- Ouborg, N. J., Y. Piquot, and J. M. Van Groenendael. 1999. Population genetics, molecular markers and the study of dispersal in plants. *Journal of Ecology* **87**:551–568.
- Ouborg, N. J., P. Vergeer, and C. Mix. 2006. The rough edges of the conservation genetics paradigm for plants. *Journal of Ecology* **94**:1233–1248.
- Ouborg, N. J., C. Pertoldi, V. Loeschcke, R. Bijlsma, and P. W. Hedrick. 2010a. Conservation genetics in transition to conservation genomics. *Trends in Genetics* **26**:177–187.
- Ouborg, N. J., F. Angeloni, and P. Vergeer. 2010b. An essay on the necessity and feasibility of conservation genomics. *Conservation Genetics* **11**:643–653.
- Parchman, T. L., K. S. Geist, J. A. Grahnen, C. W. Benkman, and C. A. Buerkle. 2010. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics* **11**:180.
- Pavli, O. I., G. I. Kelaidi, A. P. Tampakaki, and G. N. Skaracis. 2011. The *hrpZ* gene of *Pseudomonas syringae* pv. *phaseolicola* enhances resistance to rhizomania disease in transgenic *Nicotiana benthamiana* and sugar beet. *PLoS ONE* **6**:e17306.
- Pedersen, K. S., T. N. Kristensen, and V. Loeschcke. 2005. Effects of inbreeding and rate of inbreeding in *Drosophila melanogaster* – Hsp70 expression and fitness. *Journal of Evolutionary Biology* **18**:756–762.
- Pedersen, K. S., T. N. Kristensen, V. Loeschcke, B. O. Petersen, J. O. Duus, N. C. Nielsen, and A. Malmendal. 2008. Metabonomic signatures of inbreeding at benign and stressful temperatures in *Drosophila melanogaster*. *Genetics* **180**:1233–1243.
- Primmer, C. R. 2009. From conservation genetics to conservation genomics. *Annals of the New York Academy of Sciences* **1162**:357–368.
- Prins, A., J. M. Mukubi, T. K. Pellny, P. J. Verrier, G. Beyene, M. S. Lopes, K. Emami *et al.* 2011. Acclimation to high CO₂ in maize is related to water status and dependent on leaf rank. *Plant Cell and Environment* **34**:314–331.
- Ralls, K., and J. D. Ballou. 2004. Genetic status and management of California condors. *The Condor* **106**:215–228.
- Ralls, K., J. D. Ballou, B. A. Rideout, and R. Frankham. 2000. Genetic management of chondrodystrophy in California condors. *Animal Conservation* **3**:145–153.
- Romanov, M. N., M. Koriabine, M. Nefedov, P. J. de Jong, and O. A. Ryder. 2006. Construction of a California condor BAC library and first-generation chicken-condor comparative physical map as an endangered species conservation genomics resource. *Genomics* **88**:711–718.
- Romualdi, C., S. Bortoluzzi, and G. Danieli. 2001. Detecting differentially expressed genes in multiple tag sampling experiments: comparative evaluation of statistical tests. *Human Molecular Genetics* **10**:2133–2141.
- Rowe, H. C., S. Renaut, and A. Guggisberg. 2011. RAD in the realm of next-generation sequencing technologies. *Molecular Ecology* **20**:3499–3502.
- Ruijter, J. M., A. H. C. V. Kampen, and F. Baas. 2002. Statistical evaluation of SAGE libraries: consequences for experimental design. *Physiological Genomics* **11**:37–44.
- Sanchez, C. C., T. P. Smith, R. T. Wiedmann, R. L. Vallejo, M. Salem, J. Yao, and C. E. Rexroad. 2009. Single nucleotids polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. *BMC Genomics* **10**:559.
- Santure, A. W., J. Stapley, A. D. Ball, T. R. Birkhead, T. Burke, and J. Slate. 2010. On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Molecular Ecology* **19**:1439–1451.
- Slate, J., J. Gratten, D. Beraldi, J. Stapley, M. Hale, and J. M. Pemberton. 2009. Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* **136**:97–107.
- Stapley, J., J. Reger, P. Feulner, C. Smadja, J. Galindo, R. Ekblom, C. Bennison *et al.* 2010. Adaptation genomics: the next generation. *Trends in Ecology & Evolution* **25**:705–712.
- Storz, J. F., B. A. Payseur, and M. W. Nachman. 2005. Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. *Molecular Biology and Evolution* **21**:1800–1811.
- Thurston, M. I., and D. Field. 2006. *Msatfinder*: Detection and Characterization of Microsatellites. Oxford. <http://www.genomics.ceh.ac.uk/msatfinder/> (accessed on 15 October 2011).
- Tian, F., P. J. Bradbury, P. J. Brown, H. Hung, Q. Sun, S. Flint-Garcia, T. R. Rocheford *et al.* 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature Genetics* **43**:159–162.
- Tsui, C. K. M., N. Feau, C. E. Ritland, S. Massoumi-Alamouti, S. DiGuistini, L. Khadempour, and J. Bohlmann. 2009. Characterization of microsatellite loci in the fungus, *Grossmannia clavigera*, a pine pathogen associated with the mountain pine beetle. *Molecular Ecology Resources* **9**:1500–1503.
- Turner, T. L., E. C. Bourne, E. J. Von Wettberg, T. T. Hu, and S. V. Nuzhdin. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics* **42**:260–263.
- Tymchuk, W., P. T. O' Reilly, J. Bittman, D. MacDonald, and P. Schulte. 2010. Conservation genomics of Atlantic salmon: variation in gene expression between and within regions of the Bay of Fundy. *Molecular Ecology* **19**:1842–1859.
- Vera, J. C., C. W. Wheat, H. W. Fescemyer, M. J. Frilander, D. L. Crawford, I. Hanski, and J. H. Marden. 2008. Rapid transcriptome characterization for a non model organism using 454 pyrosequencing. *Molecular Ecology* **17**:1636–1647.
- Visscher, P. M., W. G. Hill, and N. R. Wray. 2008. Heritability in the genomic era – concepts and misconceptions. *Nature* **9**:255–266.
- Wall, J. D., and J. K. Pritchard. 2003. Haplotype blocks and linkage disequilibrium in the human genome. *Nature Reviews. Genetics* **4**:587–597.
- Wang, Z., M. Gerstein, and M. Snyder. 2008. RNA-seq: a revolutionary tool for transcriptomics. *Nature Reviews. Genetics* **10**:57–63.
- Wheat, C. W., H. W. Fescemyer, J. Kvist, E. Tas, J. C. Vera, M. J. Frilander, I. Hanski *et al.* 2011. Functional genomics of life history variation in a butterfly metapopulation. *Molecular Ecology* **20**:1813–1828.
- Whitlock, M. C., and F. Guillaume. 2009. Testing for spatially divergent selection: comparing Q_{ST} to F_{ST}. *Genetics* **183**:1055–1063.
- Wilhelm, B. T., and J. R. Landry. 2009. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* **48**:249–257.
- Wright, S. 1984. *Evolution and the Genetics of Populations*. University of Chicago Press, Chicago, IL.
- Zhang, J., R. Chiodini, A. Badr, and G. Zhang. 2011. The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics* **38**:95–109.