Computational and Structural Biotechnology Journal 13 (2015) 370–375

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Mini Review

Comparative genomics for biodiversity conservation

Catherine E. Grueber *

Faculty of Veterinary Science, University of Sydney, NSW, Australia
San Diego Zoo Global, San Diego, CA, USA

ARTICLE INFO

Article history:
Received 29 January 2015
Received in revised form 13 May 2015
Accepted 15 May 2015
Available online 21 May 2015

Keywords:
Adaptation
Disease
Hybridisation
Management units
Phylogeny
Threatened species

ABSTRACT

Genomic approaches are gathering momentum in biology and emerging opportunities lie in the creative use of comparative molecular methods for revealing the processes that influence diversity of wildlife. However, few comparative genomic studies are performed with explicit and specific objectives to aid conservation of wild populations. Here I provide a brief overview of comparative genomic approaches that offer specific benefits to biodiversity conservation. Because conservation examples are few, I draw on research from other areas to demonstrate how comparing genomic data across taxa may be used to inform the characterisation of conservation units and studies of hybridisation, as well as studies that provide conservation outcomes from a better understanding of the drivers of divergence. A comparative approach can also provide valuable insight into the threatening processes that impact rare species, such as emerging diseases and their management in conservation. In addition to these opportunities, I note areas where additional research is warranted. Overall, comparing and contrasting the genomic composition of threatened and other species provide several useful tools for helping to preserve the molecular biodiversity of the global ecosystem.

Contents

1. Introduction

Conservation genetics has entered the world of genomics [1]. The number of species with whole-genome sequence data is continually growing [2,3], so that more and more endangered taxa are becoming "genome-enabled" [4], that is, genome resources are available for them or their close relatives. These new technologies provide researchers with unprecedented levels of data to generate precise estimates of essential population genetic parameters, to examine questions such as the causes and genetic consequences of population decline and fragmentation [1,5; for critique see 6]. Most applied conservation genetics research targets issues operating within or amongst populations of the same species (which may be spatially or temporally separated) [7]. This level of focus is often appropriate because anthropogenic threatening processes typically occur over relatively short evolutionary time frames: the scale relevant to population/species-level processes rather than deeper evolutionary trajectories such as speciation. Nevertheless, there is additional insight to be gained from considering the evolutionary context of threatened species, i.e. by taking a comparative approach across taxa. For example, comparative analyses of species' demographic and life history characteristics have revealed those particular ecological traits that predispose species to high risk of extinction [8,9].

* Faculty of Veterinary Science, University of Sydney, NSW, Australia.
E-mail address: catherine.grueber@sydney.edu.au.

http://dx.doi.org/10.1016/j.csbj.2015.05.003

In this paper, I explore how comparative approaches using genomic data may also add value to conservation efforts.

Comparative genomics benefits most from high-quality, annotated and mapped genome data, but the pre-existence of such complete data is not necessarily a prerequisite for taking a comparative genomic approach to wildlife genetics [10]. This is good news for conservation scientists, who frequently work on non-model species for which genome resources do not exist. Several options are available, although not all will be uniformly applicable across contexts: reduced-representation libraries [11] provide a cost-effective means of obtaining genome-level data for comparative studies [12]. RNAseq to obtain transcriptome-level data can also provide valuable insight, without necessarily obtaining whole-genome data (e.g. comparative RNA sequencing of 12 primate species, most of which had little or no genomic resources [13]). Genome-informed SNP arrays, developed for well-studied species, can be used to generate large amounts of data for closely related threatened taxa (e.g. utilising a primarily domestic dog SNP array to study wild canids [14]), although the percentage of shared polymorphisms between species decays exponentially with divergence time, decreasing the amount of data obtained from the chip for more distant species [15]. A further approach to preliminary comparative genomics investigation is the generation of large amounts of sequencing data, which is then aligned to the annotated genome of a closely related species (e.g. aligning California condor sequencing data against the chicken genome [16]).

Wildlife genomics may be undertaken at multiple levels, from comparing individuals within a population (in a population genetics/genomics framework) to comparisons at higher taxonomic levels (comparative genomics). Many genomics techniques offer opportunities for conservation (for recent overviews, see [1,3,6,17]). However, despite their potential value, comparative genomic studies with explicit and specific conservation applications remain uncommon ([1,6,18], exceptions are [19,20]). Impediments to the uptake of genomics in conservation include sampling and analysis constraints [21], as well as a lack of clear examples of successful application [6], amongst others. In this review, I focus on possible applications of comparative genomics to conservation, and provide examples of a variety of avenues for future work in this field. Comparative genomics itself is a broad field, with the potential to answer many salient questions in evolutionary biology, medicine, and other fields (e.g. [22]), and therefore the analyses mentioned herein also have many applications beyond threatened species management. In fact, due to the scarcity of conservation examples, much of the empirical work I discuss here has been conducted on non-threatened species. I touch on a number of topics in brief: my aim is not to provide an exhaustive survey, but rather an overview of new ways that an ever-growing resource of genomic data can be exploited to address timely problems in biodiversity conservation.

1.1. Applications of comparative genomics to conservation

My main discussion centres on a summary of conservation science research questions that may be approached or supported by the use of comparative genomic methods, and identification of research needs to further progress these aims.

1.1.1. Characterisation of conservation units

Identifying units of conservation is a fundamental goal of any conservation strategy, essential to both resource planning in a legal and financial sense (e.g. how to distribute conservation effort) and management planning in a practical and biological sense (e.g. which populations may be mixed and which show important distinctiveness that should be preserved). Although definitions vary [23], the concept of conservation management units encompasses groupings beyond traditional taxonomic demarcations, such as evolutionarily significant units and/or variants with particular ecological or social value. Nevertheless, conservation units are usually informed by phylogeny, traditionally using putatively neutral genetic regions such as microsatellite markers or mtDNA. Importantly, these methods inform conservationists as to the degree of migration amongst putative conservation units [24] providing a distinction between "evolutionarily significant units": populations that are phylogenetically discrete, and "management units": populations with significant divergence in allele frequencies [25].

Recently, researchers have begun to target adaptive molecular variation for inclusion in the assessment of conservation units. These data introduce information about evolutionary distinctiveness into the definition of protected populations [26]. For example, diversity and differentiation at the major-histocompatibility complex (MHC), genes associated with adaptive immunity [27], have been incorporated into the delineation of conservation management units for several species, such as giant panda *Ailuropoda melanoleuca* [28] and marbled murrelets *Brachyramphus marmoratus* (a threatened seabird) [29]. However, basing management decisions on a small number of functional genomic regions presents a high risk of failing to detect evolutionarily and ecologically important processes that influence other parts of the genome [30]. Recent studies have shown how genome-level data can provide very high resolution for the reconstruction of phylogenetic trees, enabling detailed identification of species boundaries and relationships [12,31]. For example, Wagner et al. [31] recently used reduced-representation RAD sequencing to generate exceptionally detailed phylogenetic inference amongst 16 cichlid species in Lake Victoria, a community well-studied in evolutionary ecology. Evolutionary relationships amongst these species had previously been difficult to dissect using traditional methods, due to very recent divergence times which impaired discrimination amongst morphologically distinct species using much smaller numbers of nuclear and mitochondrial DNA sequence variants [31].

In conservation, phylogenetic approaches have been used to identify the most evolutionarily distinct species, which may then be targeted for particular conservation effort (e.g. EDGE [evolutionarily distinct, globally endangered] species [32]). Taking a whole-genome comparative approach to the characterisation of conservation units provides at least three advantages over traditional approaches: 1) greater resolution via the use of many more loci, 2) the ability to incorporate a wide diversity of putatively functional genetic regions (i.e. genic sequences) and 3) the ability to perform analyses using either neutral or functional data (or both), enabling researchers to study how different processes drive population structure [33]. Several challenges exist with the use of whole-genome data for the reconstruction of phylogenetic trees, such as how to conduct inference regarding species trees in the case of conflicting gene trees from different genomic regions [34]. These issues apply to all studies that use multigene data for phylogenetics, not just those with conservation aims, and their resolution is still an area of active research (e.g. [34,35]). Nevertheless, genome-level data enables researchers to determine whether any differentiation observed amongst populations results from evolutionary or demographic processes. For example, genetic structures based on different genomic regions (such as microsatellites versus MHC) are frequently uncorrelated (e.g. [29]), typically interpreted as a greater role of selection than drift at immunogenetic versus neutral loci, respectively [36]. Populations differing as a result of recent, drift-associated processes are not considered as distinct as populations differing as a result of deep adaptation processes [23]. Differentiating these mechanisms of structure amongst populations is essential to the fully informed preservation and management of molecular biodiversity in ecosystems [1].

1.1.2. Informing the conservation consequences of hybridisation

Human landscape modification has increased the frequency with which hybridisation influences the evolutionary course of many species around the world [37]. Introgression of a threatened species by a previously geographically separated and more-common relative can affect species integrity and result in extinction of the rarer type [38,39].