TITLE: Progressing in Genomic tools after population genetics approaches: Transfer of knowledge for current and future zooplanktonic organism studies

ESF Research Networking Programme

Conservation Genomics: amalgamation of conservation genetics and ecological and evolutionary genomics (ConGenOmics)

Final report of the Exchange visit (Ref. Num.: 3646)

This final report is composed of the following elements:

1. A scientific report in .PDF format, containing the following information (not exceed 6-8 A4 pages):

1.1. Purpose of the visit

As it is indicated in the application form, the current proposal for the ESF Exchange visit grant 2011 was awarded to facilitate:

A) Transfer of knowledge and introduction of the applicant to genomic tools (i.e., genome-wide markers and Illumina technology), and **B)** Preparation of future collaborations (e.g., Marie Curie-IEF grant) with a European group to share the knowledge obtained in the University of Oklahoma (please, note that the applicant discussed his impending incorporation to a Marie Curie-IOF grant to the University of Oklahoma). All these issues will benefit to the applicant currently and in the future, and will allow sharing genomic resources and knowledge between European labs and other countries.

1.2. Description of the work carried out during the visit

During the Exchange visit (five weeks in total), both the host researcher (Dr. Africa Gómez) and the applicant have been discussing and meeting to transfer of knowledge related to genomic tools in aquatic invertebrates. Dr. Africa Gómez, and her research team, is currently involved in a project using Restriction site Associated DNA (RAD) Sequencing method in the aquatic invertebrate *Triops cancriformis*. This is an innovative tool to acquire DNA sequence polymorphisms in organisms for which genomes are not yet available. The applicant discussed several technical and practical aspects related to the application of this technique with the host researcher and her lab members with the purpose to applying it in the future in research projects (please, see section *1.4*).

A) Transfer of knowledge and introduction of the applicant to genomic tools (i.e., genome-wide markers and Illumina):

RAD sequencing (or RAD tag, or RAD mapping) technology, in short:

Methods

Laboratory method

Genomic DNA is isolated from individual samples using DNA extraction kits in order to obtain in high quality and concentration (≅ 25 ng/µl), and then digested with a specifically chosen restriction enzyme (usually Sbf1 – [1]). The digest is then ligated to a double-strand adapter (*P1-adapter*), which contains 'Forward Amplification' and 'Illumina Sequencing' primer sites, as well as a 'Barcode' sequence (4-5 bp in length) for sample identification (see Fig. 1.A). To reduce erroneous sample assignment due to sequencing errors, all barcodes should differ by at least two nucleotides. The fragments ligated to different *P1-adapters* are subsequently pooled, randomly sheared (e.g., through a sonicator) and size-selected, and then ligated to a second adapter (*P2-adapter*) with the peculiarity that it is a divergent "Y" adapter containing the reverse complement of the P2 reverse amplification primer site preventing, in such a way, the amplification of genomic selected fragments lacking the *P1-adapter* (i.e., those fragments containing exclusively the *P2-adapter* ligated to each extreme). Only size-selected DNA fragments having both *P1* and *P2-adapters* will be robustly enriched by PCR amplification (Fig. 1.B).

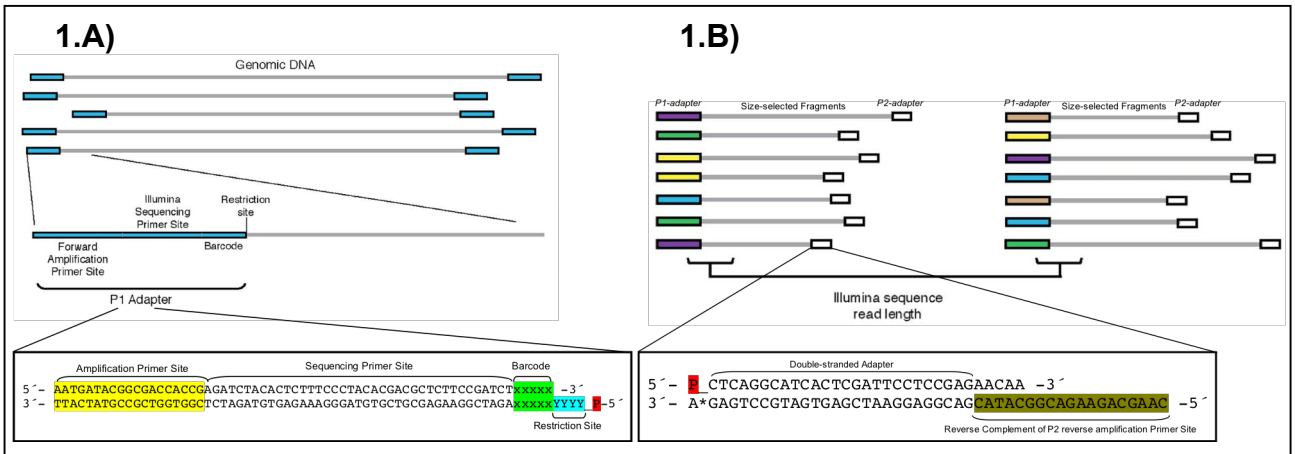

Figure 1. Schematic representation of the RAD sequencing library generation modified from [1]. **A)** Shows genomic DNA digested with a restriction enzyme and two *P1-adapters* ligated to the fragments. Particular composition of the *P1-adapter* is showed as well; **B)** Shows size-selected DNA fragments having both *P1* and *P2-adapters* ligated. These selected fragments will be selectively and robustly enriched by PCR amplification (see PCR primer combination below). Colour boxes represent *P1-adapters* with different barcodes.

The RAD sequencing library will be ready for the high-throughput Illumina Genome Analyzer Sequencing Machine (Illumina GAIIx, operating with a flow cell easily loaded) with a final selected fragment amplification and enrichment performed with P1-Forward and P2-Reverse primers (note that colours correspond with those showed in Fig. 1):

P1-forward primer: 5´- AATGATACGGCGACCACCG*A -3´

P2-reverse primer: 5´- CAAGCAGAAGACGGCATACG*A -3´

For instance, Illumina GAIIx sequencing machine (operating with 6-8 lanes) can generate 10-20x10^6 reads per lane, giving 100 bp per read (or 200 bp per paired end read), which means around 4 Gb per lane.

Data analysis methods

Currently, there are two main packages/software for RAD sequencing analyses, **RADtools** [2] and **Stacks** [3].

- **RADtools package** processes RAD sequencing data for transforming Illumina reads into candidate genetic markers. It contains four core tools. They are: RADpools, RADtags, RADmarkers, and RADMIDs. RADpools takes as input raw Illumina reads and a list of pools, and creates a directory containing a set of files, one per pool. Each file contains the reads for that pool. It separates raw reads into **pools** according to the Multiplex IDentifier (MID) sequences given to each individual sample in the RAD sequencing library. It outputs all qualities in Sanger format. RADtags takes as input the directory of files created by RADpools, and clusters the reads for each pool into candidate **tags** for that pool. RADmarkers takes as input the set of files produced by RADtags generating **candidate loci** with **alleles** (i.e., a particular tagged sequence), calling SNPs and assessing tag qualities in the process across all pools. It also looks for mismatches to tag sequences. The fourth tool, RADMIDs can be used to design and generate a set of **MIDs** of a certain length with a certain number of differences between each pair of MIDs, for use in RAD adapters. MIDs are generated with several differences between any pair so that they can be distinguished bioinformatically even if the MID for a particular read contains a sequencing error.

- **Stacks software** uses short-read sequence data to identify and genotype loci in a set of individuals either *de novo* or by comparison to a reference genome. From RAD tags sequence data, it can recover thousands of Single Nucleotide Polymorphism (SNPs) markers useful for the genetic analysis of populations. Stacks can generate markers for ultra-dense genetic linkage maps, facilitate the examination of population phylogeography, and help in reference genome assembly. In addition, Stacks makes Genome-Wide Association Studies (GWAS) more tractable in non-model species because the enormous linkage map provides a framework for the analysis of population genomic data.

Potential applications

RAD sequencing was developed for several genetic approaches such as genotyping, linkage mapping, genome scaffolding, and population genomics (e.g., [2]). However, its utility has been recently demonstrated for these and other ecological and evolutionary approaches. So, it

has been reported the parallel adaptation and evolution in the fish threespine stickleback, *Gasterosteus aculeatus* [4]; the postglacial phylogeography in the in the pitcher plant mosquito, *Wyeomyia smithii* [5]; or new investigations on the economically important plant pathogen *Phytophthora capsaci* allowing the assessment of population structure with a high recombination rate in U.S.A. [6].

Application to aquatic invertebrates

The application of genomic techniques to study the evolutionary ecology of continental aquatic invertebrates is still at a very early stage. The host researcher for this Exchange visit is currently the principal investigator in a project where RAD sequencing will be used to create a linkage map of reproductive mode for the branchiopod crustacean *Triops cancriformis*. This interesting organism shows a variation of reproductive mode between populations, meaning that dioecious, hermaphroditic and androdioecious populations can all be found across its distribution range. The resulting findings will be used to infer the genetic basis of spatial reproductive mode variation.

The applicant and the host researcher plan to apply these tools in ongoing and future projects (i.e., Marie Curie – IOF awarded to the applicant to enjoy in the University of Oklahoma, U.S.A.; and the draft of the European project Marie Curie – IEF to be applied and to be carried out in the University of Hull, U.K.). Particularly, genomic tools will be developed in two model organisms, *Daphnia* (Cladocera) for the ongoing IOF project and *Artemia* (Anostraca) for the programmed IEF project (please, read below).

1.3. Description of the main results obtained

Results obtained in the planned grant are in concordance with the objectives and the work indicated in the initial application. The main results have been the transference of knowledge regarding to genomic tools and the preparation of a draft project to be submitted in the next Marie Curie – IEF application (please, see next section *1.4*).

1.4. Future collaboration with host institution

As a second purpose indicated for the Exchange visit [i.e., **B)** Preparation of future collaborations, such as a Marie Curie IEF grant, with a European group to share the knowledge obtained in Oklahoma], here we describe a draft of the future project that will be submitted in the Marie Curie IEF application of 2013:

Future project – Marie Curie (IEF)

- Tentative title: ‘The advantage to be an invasive species: A genomic approach for continental aquatic invertebrates’

- Background:

* Model organism:

Artemia franciscana is the sexual brine shrimp most widely distributed around the world. Its native range occupies the whole American continent, occurring from Canada to temperate Chile and many Atlantic islands, but Alaska and North East of U.S.A. Previous genetic analyses have indicated the presence of substantial genetic and morphological differentiation, and high levels of local adaptation to habitat chemistry. Indeed, effective reproductive isolation between some *A. franciscana* populations could occur due to different tolerance range to ionic compositions. Additionally, *A. franciscana* cysts – harvested mainly from populations in San Francisco Bay saltworks and the Great Salt Lake in the U.S.A. – has been used worldwide as food source in aquaculture and in the pet fish trade for over half a century. In this scenario, effluents are likely to contain cysts that can potentially colonise nearby wetlands. In addition, the intentional introduction of foreign strains was promoted worldwide to increase salt production or to promote local sources of cysts. As a result of such accidental and/or intentional inoculations, *A. franciscana* has become an invasive species in saline and hypersaline wetlands worldwide. This species has been described as a 'good invasive aquatic species' with high adaptive potential (see Table and [8]) displaying different degrees of the following attributes: (1) abundance in their original range or large native range, (2) polyphagous or eurytrophic (i.e., wide dietary niche), (3) high genetic variability or phenotypic plasticity, (4) short generation times, (5) subitaneous eggs, (6) larger than most related species, (8) high dispersal rate, (9) associated with human activities (i.e., human commercialism) and (10) able to function in a wide range of physical conditions.

Biological attributes of *Artemia franciscana* contributing to its invasive success

Characteristic	Citation
Human commensal (solar saltfields, aquaculture and aquarium trade)	(Jones *et al.* 1981; Treece 2000)
Dispersal of cysts via waterbirds	(Persoone and Sorgeloos 1980; Green *et al.* 2005)
Dispersal of cysts via wind	(Persoone and Sorgeloos 1980)
Broad native geographic range	(Bowen *et al.* 1985; Van Stappen 2002)
High genetic variability	(Abreu-Grobois and Beardmore 1982; Gajardo *et al.* 1995; Torrentera and Abreu-Grobois 2002)
High concentration of phenotypic plasticity	(Browne and Wanigasekera 2000)
Euryhaline and eurythermal	(Browne and Wanigasekera 2000)
Euryoxybiont	(Persoone and Sorgeloos 1980)
Cysts have high tolerance to ultraviolet radiation and high temperature	(Triantaphyllidis *et al.* 1994; Tanguay *et al.* 2004)
Two reproductive strategies – ovoviviparity (nauplii) and oviparity (cysts)	(Criel and Macrae 2002*b*)
Sedimentary cyst bank – cysts can remain viable for over 300 years	(Marcarelli *et al.* 2005)
Ability to filter feed over a wide range of particle sizes (6.8–27.5 μm)	(Gelabert Fernandez 2001)

Table from [7]

Since we have an extensive background on this species, (see additionally the results obtained in the current manuscript in preparation indicated in section *1.5*), we propose to apply for an

IEF-Marie Curie to assess those key genes/loci involved in the capacity of invasiveness for *Artemia franciscana*. We hypothesize that those few genes involved in adaptation will evolve quickly and will be selected during the introduction events. Their local adaptation and selection will be able even to replace and remove native *Artemia* species (Mediterranean wetlands – [8]).

- Material and Methods
    * New genomic tools to be applied to this model organism.
    * RAD sequencing method depending on the genome coverage and funding acquisition.
    * THREE DIFFERENT ENVIRONMENTS FROM BOTH NATIVE AND INVASIVE GEOGRAPHIC RANGES:
    - *Artemia franciscana* from GSL and SFB ---- Native and commercialized populations.
    - *Artemia monica* -------- With specific habitat (Bowen et al 1988)
    - *Artemia franciscana* from Mediterranean invaded wetlands ---------- established invasive populations.
- Partnership with Hull University

1.5. Projected publications / articles resulting or to result from the grant (ESF must be acknowledged in publications resulting from the grantee's work in relation with the grant)
This grant has enabled the production of an article dealing with the evolutionary ecology to a continental scale of the populations of brine shrimp (Branchiopoda:Anostraca) *Artemia franciscana* (Tentative Title: 'Phylogeography of the invasive brine shrimp *Artemia franciscana* in its native American range: a role for bird migration?'). The main results obtained indicate that the American brine shrimp populations are highly subdivided supporting the existence of at least 12 divergent genetic lineages with a strong association between genetic variation and geography (i.e., isolation-by-distance pattern), but mostly regionally restricted in the American continent. The main genetic clades correspond to Pacific, Canadian and Atlantic populations mirroring shorebird migratory Americas flyways. Furthermore, it is noteworthy that these results have been those who have brought new ideas to develop in the future project to be submitted to the European Union Marie Curie – IEF grants. If these results became published, ESF would be acknowledged in the scientific paper in relation with the grant.

References cited

[1] Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3(10): e3376.
[2] Baxter SW, Davey JW, Johnston JS, Shelton AM, Heckel DG, Jiggins CD, Blaxter ML (2011) Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS ONE* 6(4): e19315.
[3] Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) *Stacks*: Building and genotyping loci de novo from short-read sequences. *G3* 1: 171-182.
[4] Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS GENETICS* 6(2): e1000862.
[5] Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, Holzapfel CM (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences* 107(37): 16196-16200.
[6] Rowe HC, Renaut S, Guggisberg A (2011) RAD in the realm of next-generation sequencing technologies. *Molecular Ecology* 20: 3499–3502.
[7] Ruebhart DR, Cock IW, Shaw GR (2008) Invasive character of the brine shrimp *Artemia franciscana* Kellogg 1906 (Branchiopoda: Anostraca) and its potential impact on Australian inland hypersaline waters. *Marine and Freshwater Research* 59: 587–595.
[8] Amat F, Hontoria F, Ruiz O, Green AJ, Sánchez MI, Figuerola J, Hortas F (2005) The American brine shrimp as an exotic invasive species in the western Mediterranean. *Biological Invassions* 7: 37-47.

2. A signed host statement form*

3. A completed and signed balance payment form* accompanied by the original travel tickets

* Items automatically generated once the online report has been submitted.