



ELSEVIER

Contents lists available at [SciVerse ScienceDirect](http://SciVerse.ScienceDirect.com)

## Comptes Rendus Biologies

[www.sciencedirect.com](http://www.sciencedirect.com)

Review/Revue

## A bird's-eye view on the modern genetics workflow and its potential applicability to the locust problem



Mohammed Bakkali

Departamento de Genética, Facultad de Ciencias, Universidad de Granada, Fuentenueva S/N, 18071 Granada, Spain

## ARTICLE INFO

## Article history:

Received 10 June 2013

Accepted after revision 23 July 2013

Available online 22 August 2013

## Keywords:

Orthoptera

Locusts

Grasshoppers

Genetics

Gene expression

Next generation sequencing

Microarrays

RNAi

## ABSTRACT

Genetics is an immense science and the current developments in its methods and techniques as well as the fast emerging tools make it one of the most powerful biological sciences. Indeed, from taxonomy and ecology to physiology and molecular biology, every biological science makes use of genetics techniques and methods at one time or another. In fact, development in genetics is such that it is now possible to characterize and analyze the expression of the whole set of genes of virtually every living organism, even if it is a non-model one. Locusts are notorious for the damage they cause to the ecosystems and economies of the areas affected by their recurrent population outbreaks. To prevent and deal with these outbreaks, we now count on both biological as well as chemical agents that are proving to be successful in reducing the damage that otherwise locust population outbreaks might cause. However, a better, efficient and environmentally friendly solution is still a hoped-for target. In my opinion, the ideal future pesticide should be both environmentally friendly, risk free and species-specific. To reach the knowledge needed for the development of such species-specific anti-locust agent, deep and accurate knowledge of the locusts' genetics and molecular biology is a must. Since genes and their expression levels lie at the bottom of every biological phenomenon, any species-specific solution to the locust problem requires a good knowledge of these organisms' genes as well as the quantitative and spatio-temporal dynamics of their expression. To reach such knowledge, collaborative work is needed as well as a clear workflow that, given the fast development in the genetics tools, is not always clear to all research groups. For this reason, here I describe a genetics workflow that should allow taking advantage of the most recent genetics tools and techniques to answer question relating to locust biology. My hope is that the adoption of this and other work strategies by different research groups, especially when the work is a collaborative one, would provide precious information on the biology and the biological phenomena that these economically important organisms exhibit.

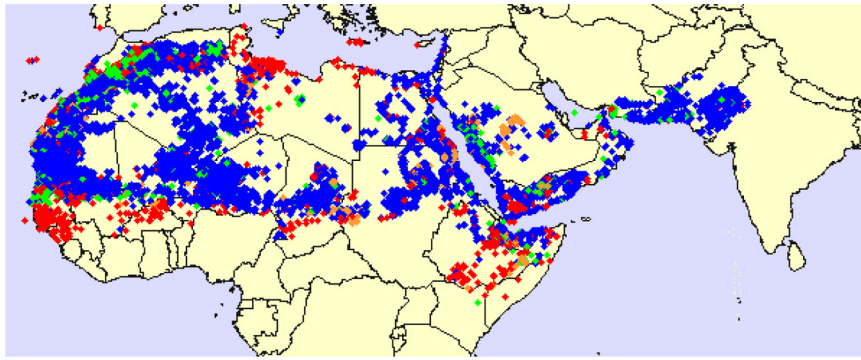
© 2013 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

## 1. Introduction

Among orthopterans, locusts are notorious for their population outbreaks and swarming capabilities and no continent is spared from locust outbreaks. Indeed, infamous among the species known to experience population outbreaks, we find *Schistocerca gregaria* and

*Oedaleus senegalensis* in Africa and parts of Asia, *Locusta migratoria* and *Oedaleus asiaticus* in Africa and Asia, *Melanoplus sanguinipes* in North America, other *Schistocerca* species (such as *S. cancellata* and *S. americana*) in South America and *Chortoicetes terminifera* in Australia. At the outbreak phase, each of these species can cover large geographic areas (Fig. 1), often with disastrous consequences. Indeed, millions of dollars are spent each year in order to control the population size of such species that, otherwise, might experience outbreaks and become even

Email address: [mbakkali@ugr.es](mailto:mbakkali@ugr.es).



**Fig. 1.** Geographical distribution of *Schistocerca gregaria* populations sighted between 01-01-2000 and 01-01-2013. Green dots: solitary hoppers; blue dots: solitary adults; orange dots: gregarious hoppers; red dots: gregarious adults. United Nations Food and Agriculture Agency (<http://www.fao.org/ag/locusts/en/info/info/index.html>).

more costly. For instance, a document on the benefit-cost analysis of locust control operations for 2010–2011 by the Australian Plague Locust Commission (<http://www.daff.gov.au/animal-plant-health/locusts>) states: “Locust control operations this season are estimated to have avoided potential losses of \$963 million. Total expenditure by all parties was estimated to be \$50 million. The net benefits of control are therefore \$913 million, with an estimated ratio of benefits to costs of around 19.2:1.”

Thus far, the solution to the outbreaks problem is largely based on prevention, with significant governmental personnel and resources dedicated to this task (e.g., the Moroccan anti-locust centre <http://www.criquet-maroc.ma/>, the pest research unit of the French CIRAD <http://ur-bioagresseurs.cirad.fr/>, and the Kenyan ICIPE <http://www.icipe.org/>). Coordination between countries is also needed and a section of the United Nations Food and Agriculture Organization is dedicated to this issue (see <http://www.fao.org/ag/locusts/en/info/info/index.html>). The strategies adopted for preventing locust outbreaks are proving helpful but, when they fail, the options are few. Although entomopathogenic fungi (such as *Metarhizium anisopliae* var. *acridum*) are currently available (e.g., <http://www.lubilosa.org/>) and, sometimes, used as a bio-control agents for treating the locust outbreaks, the pesticides option is still on the table and prevalent (e.g., Fipronil). No need to stress here that governments, society and local communities are increasingly aware of the dangers of the excessive use of pesticides (e.g., <http://www.panna.org/>). Neither it is necessary to extend on the issue of the non-specificity of the pesticides that, when used, affect the targeted locust as well as other organisms in the treated area, including humans, other mammals, bees... (e.g., [1–5]).

In my opinion, a species-specific treatment would be the ideal solution. For that, a deep knowledge of the molecular biology of the target species and its peculiarities is needed. Such knowledge would also offer precious data to many fundamental biological questions as well as to comparative studies (including the model organisms). Much is known about grasshoppers and locusts and the interest of researchers on this subject is continuously growing. In fact, a search for the word locust on the Pubmed database of the US National Center

for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/pubmed/>) shows strong positive correlation between the number of published papers and the year of publication ( $r = 0.871$ ). In fact, the numbers went from 25 papers published in 1970 to 221 in 2012, with 22 articles published just in January 2013 (Fig. 2a). However, these numbers are misleading, and their real meaning is only revealed when compared to the numbers from research on other organisms. For *Drosophila*, the number of published works on the NCBI database shows stronger correlation with the year of publication ( $r = 0.957$ ). It grew from 331 papers published in 1970 to 3711 in 2012, with 406 papers published just in January 2013 (Fig. 2b). Strikingly, the ratio of locust articles to *Drosophila* articles per year shows that the number of publications on locusts represents just around 5% of the number of published works on *Drosophila* (Fig. 2c). Even more, the between-years increase in the number of research papers is much more irregular when it comes to locusts than it is for *Drosophila* (Fig. 2d). One might argue that comparing with a model organism may not be the right way to assess the state of the research on locusts. That is partially right, but the word locust includes several species. In addition, each of these species represents a serious (sometimes almost existential) threat to entire regions of the globe. One would thus expect research on these organisms to represent a bit more than a meager 5% of the research on *Drosophila*. Furthermore, the number of published works on locusts in 2012, for instance, is lower than that on ants (526), bees (580), and spiders (517).

The number of published articles is directly dependent on the number of research groups and on the level of financing (the first also dependent on the second in a way that the more financing a research area, the more attractive it gets to researchers). Discussing the reasons for this near-precarious state of the research on locusts and the ways to overcome it could by itself be a theme for a manuscript and would probably be better discussed in meetings, conferences and other scientific gatherings. What matters for us now is that we are still far from reaching the level of knowledge that should allow a deep understanding of the molecular mechanisms associated with locust population outbreaks, and efficient comparative studies to determine the species-specific processes

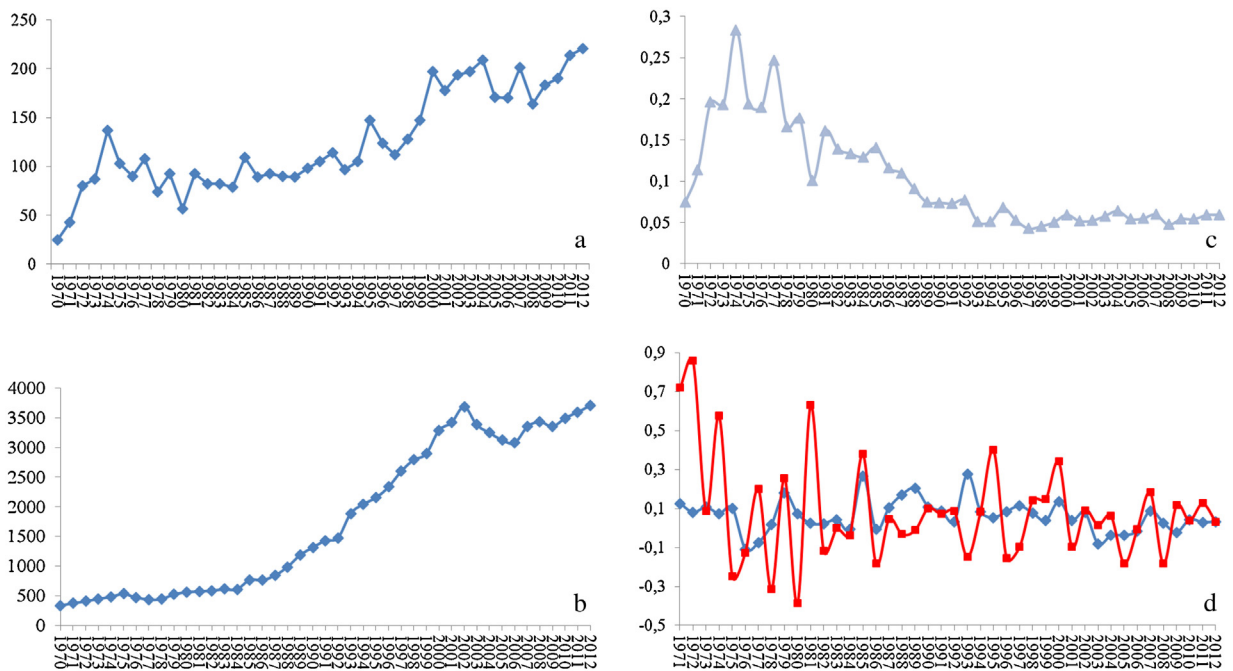


Fig. 2. Statistics of the published peer-reviewed science on locusts and *Drosophila*. (a) Number of published articles on locusts per year, (b) number of published articles on *Drosophila* per year, (c) ratio of the number published articles on locusts to that on *Drosophila* per year, and (d) inter-annual increment in the number of articles published on locusts (red line) and *Drosophila* (blue line).

and molecules that I hope might serve as target to those hopped for species-specific pesticides.

Fortunately, technological developments in the areas of molecular biology and genetics now make the possibility of a deep knowledge of the molecular aspects underlying the peculiarities of a species easier than ever. In genetics, the key words to such objective are comparative studies aimed at gene identification, gene location, analysis of gene expression, analysis of gene regulation and analysis of gene interactions and function. A battery of techniques is now available for such studies, and these include high-throughput sequencing (also called next generation or deep sequencing), *in situ* hybridization techniques, microarrays and quantitative PCRs, epigenetics study techniques and gene knockdowns. Combined together, and in the right order, these techniques should allow the characterization of the locust genes and of their functions. I therefore briefly introduce here the most relevant among these techniques as well as their potential usefulness for the locust research and a potential workflow that an average genetics laboratory could follow to get data on genes and their expression and function. My hope is that the combined research efforts would provide results that, when taken together, would provide valuable scientific knowledge and a good-enough basis for applied research on environmentally friendly and/or species-specific locust pesticides.

## 2. Genomics

Genome sequences of a number of non-locust pest species are now available and some research centers are

now engaged in deciphering even more genomes (e.g., The Insect Genome Research Unit of the Japanese Agronomics Research Center, <http://www.nias.affrc.go.jp/eng/org/Agrogenome/Insect/index.html>). Notorious examples of such genomes include those of the mosquito *Anopheles gambiae* [6], the red flour beetle *Tribolium castaneum* [7] and, recently, the pine beetle, *Dendroctonus ponderosae* [8]. However, obtaining locust DNA sequences, as we all know, means sequencing and, when it comes to that, we are faced with two main objectives: firstly, identifying the expressed sequences of the genome, and secondly, identifying the genome itself. The second question remains thus far untouched because of the difficulties that orthopteran genomes present. In fact, the genome of a species such as *S. gregaria* is two and half times the size of the human genome [9]; making any genome sequencing project too costly for the average genetics laboratory. In addition, the large size of the orthopteran genomes is due to its shared amount of repeated DNAs. From that comes the second difficulty; which is related to the challenge that represents assembling sequence reads that contain parts of the repeated DNA. Two sequencing reads of two separated genomic regions (even in different chromosomes) that share the same repeated DNA would be falsely assembled into one non-real sequence, due to the overlap between the repeated DNA parts of these reads. This issue could be surmounted only if:

- the sequence reads were long enough (larger than the largest repeated DNA unit in the genome);
- the sequencing depth were high enough as to sequence every possible genomic window of the same size as the

sequencing read (that may mean a too costly sequencing effort at hundreds fold the size of the genome);

- if the sequence assembly algorithms were able to discriminate between events of false joining of repeated DNA-containing reads and the correct assembly of genuine sequences containing the repeated elements.

For the size of the read, currently the 454 GS FLX plus system is probably the best technological option, as it offers up to 1000 bases per read (<http://454.com/products/gs-flx-system/index.asp>). Still, with just one million reads per plate, tens of plates are needed to get the desired sequencing depth, which, at a cost of around 15,000 US \$ per plate, remains a very costly option for the average research laboratory. In addition, even the best assembly algorithms (e.g., ABYSS [10]) still need to be improved to deal with the assembly ambiguities that result from the repeated DNAs. Thus, as long as the algorithms for DNA sequence assembly do not deal with the issue of repeated sequences and/or the sequencing technology becomes even more affordable and produces even larger sequences, all what we can aim for at this point is to construct some large contigs of the genome or to specifically sequence genomic regions around a specific gene. The latter objective might be of great importance, as the genomic region of a gene is what contains its *cis*-regulatory elements and, thus, studying it could be key to characterizing that gene's activators and suppressors, as well as its interactions with other genes. For such objective, a chromosome walking strategy is needed and, although the technique was originally quite lengthy and technically demanding, several commercially available kits now make it both more affordable and attainable.

Tightly linked to the availability of genomic sequences are the epigenetic analyses. These are currently approached through detection of the methylation islands and of their variation between physiological conditions (for a review, see [11]). A couple of preliminary studies have recently been carried out on *L. migratoria* [12] and recently on *S. gregaria* [13]. In the absence of complete genomes, the first work used a 'traditional' enzymatic approach and the second used the bisulfate method on incomplete genomic DNAs. The bisulfite-sequencing technology is currently the tool of choice for methylation and epigenetics studies [14,15]. This method relays on converting the non-methylated cytosine residues of the DNA to uracil, while leaving the methylated cytosines intact. Subsequent sequencing of the genomic DNAs and the comparison of the untreated sequences to the treated ones (or of the bisulfite-treated DNA sequences from two physiological conditions) allows identification of the changes in methylation patterns associated with the studied physiological phenomenon. The availability of reference genomic sequences rather than expressed transcripts is of great importance to these methylation studies; as most of the epigenetic differences between the compared conditions (e.g., gregarious vs. solitary) would be due to either silencing or modification of the level of expression of certain genes. Both effects are largely due to modifications at the *cis*-regulatory elements (promoters and enhancers), which would not be available unless

genomic DNAs are sequenced. Although the task is not easy, our lab is engaging in an initial locust genome sequencing project, whose first results should be available within the two coming years.

### 3. Transcriptomics

Contrarily to the genomic DNA, sequencing the transcripts of the expressed orthopteran genes is currently both possible and affordable. When it comes to the sequencing technology, sequencing using Illumina technology is the most cost effective and the Illumina HiSeq2000 system is probably the best among the currently available options (<http://www.illumina.com/systems/ilmn>). Although the read size is, for now, just up to 100 pb, this technique produces millions of reads (a full HiSeq2000 sequencing run produces 6 billion reads of 100 bases in 11 days). The sequencing reads are then easily filtered and assembled into expressed sequence tags (ESTs) and full transcripts. For this, sequence assembly software such as ABYSS [10], TRANSABYSS [16] VELVET [17] OASES [18] and TRINITY [19], are proving to be easy to handle and efficient in assembling expressed tag sequences from deep transcriptome sequencing. Of course, the reason for this difference between genome and transcriptome sequencing is the lack of the repeated DNA problem in the latter case, the average length of the sequences to assemble (genes vs. full chromosomes) and the absence of need for mapping and assigning sequences to chromosomes.

Depending on the sequencing depth, also called coverage (i.e., the number of times any given nucleotide of the transcriptome is expected to be sequenced), Illumina sequencing also allows comparative analyses of the gene expression levels between physiological conditions – a method referred to as RNA-seq. The logic here is that there is a positive correlation between the relative number of sequencing reads that correspond to a gene and the relative level of expression of this gene. This way, the more reads of a sequencing experiment contribute to the assembly of an EST (or transcript), the more molecules of the corresponding gene must have been present in the sequenced library (indicative of the relative number of molecules per cell). Short-sequence alignment tools such as the Burrows–Wheeler aligner (BWA) [20] allow the alignment of the raw sequencing reads to the assembled contigs (ESTs and transcripts) and algorithms such as Cufflinks [21] and DESeq [22] do the comparative counts of the alignments events between the raw sequencing reads and the assembled ESTs and transcripts.

Indeed, we have separately sequenced the transcriptome of the desert locust *S. gregaria* from different tissues of both solitary and gregarious specimens (Fig. 3), and, with an estimated  $120 \times$  sequencing depth per tissue, we have assembled the expressed tag sequences and unigenes, and we are now in the phase of the comparative analysis of the gene expression levels. With this, we expect to identify genes that show differential expression between the harmless phase of this locust and their outbreak phase. This will be the first high-throughput comparative sequencing of *S. gregaria* transcriptome. Prior to this, an impressive comparative sequencing of the solitary and





Fig. 3. Gregarious (left) and solitary (right) nymphs of the desert locust *Schistocerca gregaria* reared at our laboratory.

gregarious *S. gregaria* nervous system transcriptome was carried out using the by-now 'old fashioned' ABI sequencing technology [23]. However, the pioneering large-scale sequencing project was that of the *L. migratoria* transcriptome by Lee Kang's research team [24] – also based on ABI sequencing technology. Later, the same research group went for further sequencing of the *L. migratoria* transcriptome; this time using high-throughput technology [25].

Of course, prior to the analysis of gene expression levels, one has to annotate the assembled sequences (annotation could also be done after analysis of the expression levels, but that would be counterintuitive). For this, a recommended initial strategy is to carry out local blast searches against the annotated sequences from organisms such as *Drosophila melanogaster*, *Tribolium castaneum*, *Bombyx mori*, *Apis mellifera*, *Nasonia vitripennis*, *Aedes aegypti*... These searches could be complemented by the use of the Blast2Go software [26], which provides further automated annotation steps.

In conclusion, the sharp drop in sequencing costs and the power and efficiency of the new sequencing technologies as well as the increasing sophistication of the bioinformatics algorithms and software used for assembling, annotating and analyzing sequences and their variation (isoforms) and levels of representation in the sample are making accurate and complete sequencing of the locust transcriptomes possible. Data from these sequencing efforts will no doubt pinpoint key genes for the locust outbreak behavior and/or for specific targeting of the locust of interest.

#### 4. Characterization of gene expression

With the deep sequencing data of locust transcriptomes available, a suite of techniques for gene mapping and analysis of expression are now possible. These include *in situ* hybridization, immune-staining, quantitative (or real-time) Polymerase Chain Reactions (qPCR) and microarrays. Thinking of *in situ* hybridization means to deal with two variables: fluorescent *versus* non-fluorescent and on chromosomes *versus* on tissues. Detecting repeated DNAs on chromosomes is already routine technique for many

orthopteran research groups (e.g., [27,28]). Detecting a single-copy gene on chromosomes is possible (e.g., [29,30]) and, in a way, has been applied to locusts [31]. Still, this remains a highly challenging task that requires a sufficient sequence length and a good fluorochrome, so the adaptation and use of some of its variations, such as branched DNA *in situ* hybridization [32] or PCR *in situ* hybridization [33], might be a better option. The other *in situ* hybridization, on tissues, sections and whole mounts, uses fluorescence or non-fluorescence labeled complementary single-stranded DNA or RNA to detect the transcript of one or several given genes in the tissue or the organism in general (see [34–38]). Analysis of gene expression using *in situ* hybridization is often used locusts (e.g., [39–42]). Except from the fact that one technique physically detects the location of a gene on the corresponding chromosome whilst the other detects the cells where this gene is actually expressed, and that the former technique needs fluorescence and is more challenging, the *in situ* hybridization principle is the same and its protocol is significantly similar (for instance, compare protocols in the references herein).

With the gene sequence on hand, another possibility for gene expression analysis is to detect expression after the transcript has been translated into protein. The technique of choice here is immunofluorescence staining (e.g., [43–46]). The procedure starts by inferring the amino acids sequence from the sequenced transcript then synthesizing it and producing an antibody against the synthesized amino acids sequence (called primary antibody) in a host animal (e.g., mouse, goat). When infused into tissues and cells, this antibody should specifically bind to the protein produced by the gene of interest. A labeled second antibody (called secondary antibody), which depends on the host used for producing the primary antibody, is then used so that it can bind to and reveal the primary antibody through its labeling (fluorescent normally). Although immune-staining is a powerful technique that has been used for locusts in many occasions (e.g., [46,47]), it remains expensive and, like *in situ* hybridization on tissues and whole mounts, does not provide quantitative results. The use of this technique could thus be replaced by the more affordable *in situ* hybridization, unless a primary antibody is available and tested for specificity, or if what the researcher is looking for is to infer the site of action of the gene product (membrane localized, cytoplasmic, nuclear...). It is worth mentioning here that many of the antibodies used for *Drosophila* as well as other invertebrates are known to work for Orthoptera (e.g., [48]).

The availability of transcripts, from deep transcriptome sequencing projects, therefore provides a good set of potential sequences to either map on the chromosomes or detect and follow the evolution of their expression between tissues, stages and physiological conditions. However, if *in situ* hybridization could be done with just a single-gene (normally part of it) that might even be obtained through amplification using degenerated primers (thus no need for deep sequencing), one of the powerful techniques of gene expression analysis is microarrays [49–52] and, this necessarily needs a good set of sequences (such as those provided by deep sequencing). As stated

above, high-throughput sequencing, especially Illumina HiSeq2000, allows analysis of gene expression levels (RNA-seq). Still, these data are usually further validated through the use of microarrays. The latter comes in two flavors. The by-now 'old fashioned' on-slide microarrays that relays on the synthesis of oligonucleotides (between 60 and 70-mer), each of them complementary to one end of one transcript (generally the 3' end). Here, the oligonucleotides are printed as spots on the slide and the labeled cDNA of the transcriptome of interest are then hybridized against the printed oligomers using a similar protocol to the one used for the *in situ* hybridization of chromosomes and tissues. This technique allows comparative analysis of the relative gene expression levels between samples and conditions as one can label each DNA sample using a different fluorochrome (red *versus* green), mix the cDNAs in equal quantities and let them compete for hybridization against the spots of oligos printed on the microarray slide. This is a nice and powerful technique for gene expression analysis that has provided precious data on various research topics, but it suffers from two drawbacks: the cost of the oligos and the physical limitations of the slide that do not allow for as much sequences as a transcriptome may require. At the near cutting-edge of the technology, we now can opt either for Roche NimbleGen or Agilent's costume gene expression microarrays, the latter in its SurePrint HD or SurePrint G3 formats. With this technique, microarray chips can now take up to one million sequences and the costs are lowered down as, instead of synthesizing the oligos and printing them on the slide, one can now opt for building the oligos straight on the array chip. In addition, the great development in the optics and image analysis algorithms is making the microarrays even more sensitive. When it comes to locusts, the first, pioneering, microarray work was that of Le Kang group on *L. migratoria* [53] and, with our transcriptomic sequencing data ready, we are planning a microarray experiment for *S. gregaria*.

Another technology that allows detection and comparison of gene expression levels is the qPCR. I would not go into much detail about this technique as it is currently used almost routinely in most molecular biology and genetics laboratories. It is based on an estimation of the number of cDNA molecules corresponding to a particular transcript by real-time detection of the dynamics of a PCR amplification of its double-stranded molecules. A series of normalizations and comparison with the amplification dynamics of few control sequences (usually three house-keeping genes) thus allow estimation of the relative abundance of the molecules corresponding to our gene of interest in the cDNA library being analyzed. Detection of the number of double-stranded DNA molecules and the following of the PCR dynamics is possible thanks to the detection of fluorescence levels of a fluorochrome that emits more fluorescence when it is associated with double-stranded DNA molecules than when it is associated with single-stranded DNA molecules (SYBER Green). Although this technique is relatively old and cannot produce masses of data, as it detects sequence by sequence (we cannot usually mix different transcripts in the same reaction tube), it remains an almost necessary tool for validation of both deep sequencing and microarray data. In fact, no deep

transcriptome sequencing or microarray project would be complete without validation of some of the data (selected genes) via qPCR.

## 5. Functional genetics

With the transcriptome and/or the genomic DNA sequenced, the genes of interest selected, mapped, and their expression analyzed both in space (tissues) and in time (conditions or stages), what else can a geneticist do? The answer could be the analysis of the function of the gene.

When at this point, the researcher would already have had a large set of genes. Some of these would be identified as potentially involved in the researched phenomenon (based on the quantitative differences of their expression levels). The expression of some of these genes would also have been linked to organs, tissues or cell types. What remains is to study the effects of these genes and their potential interactions. This could be done by altering the gene of interest or its expression level. For this, one can think of two potential methods: transgenesis and gene knockdown.

The locust genome is full of transposons of all kinds. Indeed this could be suspected from the very large size of the orthopteran genomes and was, in fact, confirmed in *L. migratoria* [54,55] as well as in *Saga pedo*, *Eyprepocnemis plorans*, *Dociostaurus maroccanus*, and *S. gregaria* (Data in preparation from our multiple transcriptome sequencing projects). We also saw that transposases are highly expressed in locust cells (in preparation). While this may be an interesting fact for research on genome instability, unfortunately, it means that transgenesis would be a very challenging task as one has to find a transposon whose transposase is not naturally expressed in the locust genome. Furthermore, the large size of the locust genome means that the transposition events would be very likely to land on chromosomal regions of repeated DNAs, where the transformed sequence may not be expressed or may present inconsistent expression patterns because of the position effect variegation.

Fortunately, there is a technique that is now proven to allow altering gene expression levels. RNA interference-mediated knockdown of gene expression is increasingly used for decreasing transcript levels by up to 90% and, thus, inferring gene function and interactions from the resulting phenotypes. Knockdown of a gene expression through the use of RNA interference technique (RNAi for short) relays on the introduction into the cells of double-stranded RNA molecules that correspond to part of the target gene. Using a well-known set of enzymes (see [56–59]), the cell fragments the introduced double-stranded RNA and uses the resulting fragments as templates for fragmenting the messenger RNAs that contain such template. This normally results in the degradation of the messenger RNA of the targeted gene and, thus, the prevention of its translation and protein synthesis. The final result is hence the abolition of the target gene function and interactions (for a review on RNAi see [60,61]). Delivery of the double-stranded RNA into the cells is easier than it sounds given the presence of a cross-membrane RNA transport protein

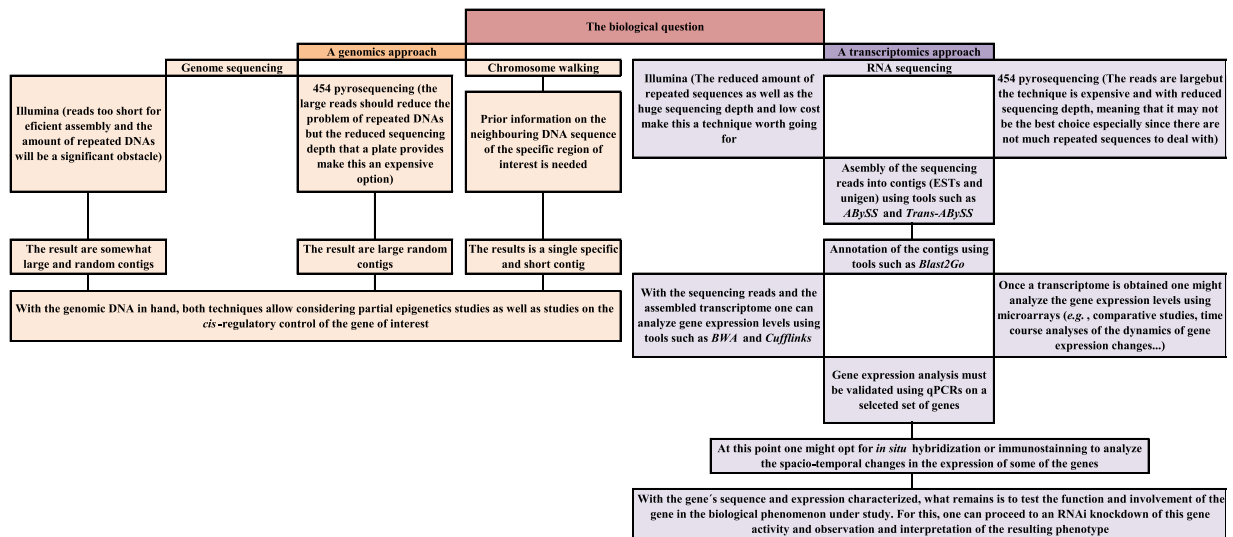


Fig. 4. Schematic summary of the modern genetics workflow suggested for a deep analysis of the genetic basis of phenomena shown by locusts (such as the capacity to shift from solitary to gregarious behaviors).

(called SID1 [62,63]). Fortunately, locust cells express this protein, as proven by the success of RNAi in grasshoppers and locusts. The pioneering RNAi work on an orthopteran species was by Dong and Friedrich [64] on *S. americana* and, since then, the technique was proven successful in other species such as *L. migratoria* [53] – in our laboratory, RNAi has proven useful in the two species where we tested it, *E. plorans* and *S. gregaria* (e.g., [65]).

If knocking down a gene's expression would allow inferring its function, a microarray study of the RNAi-treated transcriptomes should provide information on the genes lying downstream of the studied one in the gene network. Unfortunately, we are all too aware that the average laboratory cannot financially afford a microarray study after each gene knockdown experiment. So, unless one wants to go for qPCR and test individual genes whose interaction with the gene of interest was previously reported or inferred from the phenotype observed after the RNAi experiment, at this point, one can take the study of the gene of interest as completed since the gene of interest would have been identified, its expression characterized both in quantity, space and time, and its function inferred.

The use of RNAi as a tool for dealing with pest insects is debated since few years now (e.g., [66,67]) and screening for potentially useful genes to use as pesticides was already done for few species (such as Beet Armyworm, *Spodoptera exigua* [68]) and a patent on controlling pests using RNAi that includes locusts (patent number EP1934357 B1) was granted in 2011 to Els Van Bleu et al.

## 6. A final note

Genetics is a vast and fast evolving science and the tools it offers are currently used for solving questions ranging from taxonomy to physiology and molecular biology. The genetics techniques and the experimental strategies a researcher can adopt are countless and the workflow

described in this article (Fig. 4) is just one of the many possible ways one can adopt. In addition, we tend to venerate the new developments but, in my opinion, it is worth mentioning that, by no means modern genetics techniques make classical genetics techniques obsolete.

I started this essay by introducing the potentially 'applied' aspect of researching locust genetics and the reader may be left with the feeling that Orthoptera research is of no use unless it is to deal with the problems that these organisms cause. Of course, increasing the research effort on the orthopteran molecular biology and genetics would, and I expect will, provide useful information and ideas to those of us who are interested in developing alternative pesticides; which I think should be both environmentally friendly and species-specific. However, orthopteran chromosomes and genetics have been used as key material not only for scientific research (especially on chromosomes and repeated and non-coding DNAs), but also for teaching the science of genetics. Due to the large size of the chromosomes and the ease of observing them under the microscope, Orthoptera are a material of choice for the teaching of cytogenetics and meiosis. Even more, the upcoming stream of genetics and molecular biology data on locusts and other orthopteran species should serve for comparison with and validation of the scientific facts obtained from research on model organisms.

Indeed, locusts and other orthopteran species do not make it as model organisms for the genetics studies as they are 'too complex' and not as easy to rear and study as model organisms are (i.e., too mundane). But they are closer to most of the 'real' organisms than model organisms are. In fact, model organisms are chosen as such precisely because of characteristics that make them unique and less representative of most of the living beings (i.e., most of the organisms that the model one is meant to represent do not have simple traits, their genomes are not small, they take time to reproduce and do not produce as

many progeny as the model organisms do). It is my personal reflection that, with the advances in genetics and molecular biology tools, it is probably time to move from the model organism to more 'realistic' ones. Orthopteran species are both very 'realistic' and economically important (in the negative sense). They therefore have the potential to be a material of choice for the post-model organism era which I predict is about to emerge.

### Disclosure of interest

The author declares that he has no conflicts of interest concerning this article.

### Acknowledgements

I am thankful to my wife Pernille Lavgesen for her continuous understanding and support. I am also thankful to Prof. Josefa Cabrero Hurtado and Prof. Juan Pedro Martinez Camacho for their support and helpful comments on this manuscript. This work was supported by a Spanish Ramon y Cajal Fellowship and a grant from the Spanish Ministerio de Ciencia e Innovación (BFU2010-16438).

This article is dedicated to my mother, Fatima Allal Errahhaoui, and children, Nadia Lavgesen Bakkali and Elias Lavgesen Bakkali, and in memory of my father Abdeslam Bakkali (1945–2011).

### References

- [1] P. Varro, J. Gyori, I. Vilagi, In vitro effects of fipronil on neuronal excitability in mammalian and molluscan nervous systems, *Ann. Agric. Environ. Med.* 16 (1) (2009) 71–77.
- [2] C.P. Pereira, et al., Effects of fipronil (active ingredient of Frontline) on salivary gland cells of *Rhipicephalus sanguineus* females (Latreille, 1806) (Acari: Ixodidae), *Vet. Parasitol.* 166 (1–2) (2009) 124–130.
- [3] J. Leghait, et al., Fipronil-induced disruption of thyroid function in rats is mediated by increased total and free thyroxine clearances concomitantly to increased activity of hepatic enzymes, *Toxicology* 255 (1–2) (2009) 38–44.
- [4] T.L. Lassiter, et al., Is fipronil safer than chlorpyrifos? Comparative developmental neurotoxicity modeled in PC12 cells, *Brain Res. Bull.* 78 (6) (2009) 313–322.
- [5] A.K. El Hassani, et al., Glutamatergic and GABAergic effects of fipronil on olfactory learning and memory in the honeybee, *Invert. Neurosci.* 9 (2) (2009) 91–100.
- [6] R.A. Holt, et al., The genome sequence of the malaria mosquito *Anopheles gambiae*, *Science* 298 (5591) (2002) 129–149.
- [7] S. Richards, et al., The genome of the model beetle and pest *Tribolium castaneum*, *Nature* 452 (7190) (2008) 949–955.
- [8] C.I. Keeling, et al., Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major forest pest, *Genome Biol.* 14 (3) (2013) R27.
- [9] P.J. Wilmore, A.K. Brown, Molecular properties of orthopteran DNA, *Chromosoma* 51 (4) (1975) 337–345.
- [10] J.T. Simpson, et al., ABySS: a parallel assembler for short read sequence data, *Genome Res.* 19 (6) (2009) 1117–1123.
- [11] P.W. Laird, Principles and challenges of genomewide DNA methylation analysis, *Nat. Rev. Genet.* 11 (3) (2010) 191–203.
- [12] K.L. Robinson, et al., Evidence for widespread genomic methylation in the migratory locust, *Locusta migratoria* (Orthoptera: Acrididae), *PLoS One* 6 (12) (2011) e28167.
- [13] C. Falckenhayn, et al., Characterization of genome methylation patterns in the desert locust *Schistocerca gregaria*, *J. Exp. Biol.* 216 (Pt 8) (2013) 1423–1429.
- [14] J. Reinders, J. Paszkowski, Bisulfite methylation profiling of large genomes, *Epigenomics* 2 (2) (2010) 209–220.
- [15] R.P. Darst, C.E. Pardo, L. Ai, K.D. Brown, M.P. Kladdé, Bisulfite-sequencing of DNA, *Curr. Protoc. Mol. Biol.* chapter 7 (2010) (Unit 7.9.1–17).
- [16] G. Robertson, et al., De novo assembly and analysis of RNA-seq data, *Nat. Methods* 7 (11) (2010) 909–912.
- [17] D.R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs, *Genome Res.* 18 (5) (2008) 821–829.
- [18] M.H. Schulz, et al., Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels, *Bioinformatics* 28 (8) (2012) 1086–1092.
- [19] M.G. Grabherr, et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome, *Nat. Biotechnol.* 29 (7) (2011) 644–652.
- [20] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics* 25 (14) (2009) 1754–1760.
- [21] C. Trapnell, et al., Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, *Nat. Biotechnol.* 28 (5) (2010) 511–515.
- [22] S. Anders, W. Huber, Differential expression analysis for sequence count data, *Genome Biol.* 11 (10) (2010) R106.
- [23] L. Badisco, et al., Transcriptome analysis of the desert locust central nervous system: production and annotation of a *Schistocerca gregaria* EST database, *PLoS One* 6 (3) (2011) e17274.
- [24] L. Kang, et al., The analysis of large-scale gene expression correlated to the phase changes of the migratory locust, *Proc. Natl. Acad. Sci. U S A* 101 (51) (2004) 17611–17615.
- [25] S. Chen, et al., De novo analysis of transcriptome dynamics in the migratory locust during the development of phase traits, *PLoS One* 5 (12) (2010) e15633.
- [26] A. Conesa, et al., Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, *Bioinformatics* 21 (18) (2005) 3674–3676.
- [27] M. Bakkali, et al., Population differences in the expression of nucleolus organizer regions in the grasshopper *Eyprepocnemis plorans*, *Protoplasma* 217 (4) (2001) 185–190.
- [28] J. Cabrero, et al., Common origin of B chromosome variants in the grasshopper *Eyprepocnemis plorans*, *Heredity (Edinb.)* 83 (1999) 435–439.
- [29] M. Bugno, et al., Fluorescent in situ hybridization mapping of the epidermal growth factor receptor gene in donkey, *J. Anim. Breed. Genet.* 124 (3) (2007) 172–174.
- [30] J.C. Lamb, et al., Single-gene detection and karyotyping using small-target fluorescence in situ hybridization on maize somatic chromosomes, *Genetics* 175 (3) (2007) 1047–1058.
- [31] D.E. Ferrier, M. Akam, Organization of the Hox gene cluster in the grasshopper, *Schistocerca gregaria*, *Proc. Natl. Acad. Sci. U S A* 93 (23) (1996) 13024–13029.
- [32] A.N. Player, et al., Single-copy gene detection using branched DNA (bDNA) in situ hybridization, *J. Histochem. Cytochem.* 49 (5) (2001) 603–612.
- [33] E.J. Speel, F.C. Ramaekers, A.H. Hopman, PCR-based detection of nucleic acids in chromosomes, cells, and tissues: technical considerations on PRINS and in situ PCR and comparison with in situ hybridization, *Methods Mol. Biol.* 226 (2003) 405–424.
- [34] J. Asp, A. Abramsson, C. Betsholtz, Nonradioactive in situ hybridization on frozen sections and whole mounts, *Methods Mol. Biol.* 326 (2006) 89–102.
- [35] G. Hauptmann, One-, two-, and three-color whole-mount in situ hybridization to *Drosophila* embryos, *Methods* 23 (4) (2001) 359–372.
- [36] S.C. Hughes, H.M. Krause, Double labeling with fluorescence in situ hybridization in *Drosophila* whole-mount embryos, *Biotechniques* 24 (4) (1998) 530–532.
- [37] S.C. Hughes, et al., Fluorescence in situ hybridization in whole-mount *Drosophila* embryos, *Biotechniques* 20 (5) (1996) 748–750.
- [38] M.A. Nieto, K. Patel, D.G. Wilkinson, In situ hybridization analysis of chick embryos in whole mount and tissue sections, *Methods Cell Biol.* 51 (1996) 219–235.
- [39] P.K. Dearden, M. Akam, Early embryo patterning in the grasshopper, *Schistocerca gregaria*: wingless, decapentaplegic and caudal expression, *Development* 128 (18) (2001) 3435–3444.
- [40] C.C. Chang, P. Dearden, M. Akam, Germ line development in the grasshopper *Schistocerca gregaria*: vasa as a marker, *Dev. Biol.* 252 (1) (2002) 100–118.
- [41] P. Dearden, M. Akam, A role for Fringe in segment morphogenesis but not segment formation in the grasshopper, *Schistocerca gregaria*, *Dev. Genes Evol.* 210 (7) (2000) 329–336.
- [42] P. Dearden, et al., Maternal expression and early zygotic regulation of the Hox3/zen gene in the grasshopper *Schistocerca gregaria*, *Evol. Dev.* 2 (5) (2000) 261–270.
- [43] B.T. Bennett, J. Bewersdorf, K.L. Knight, Immunofluorescence imaging of DNA damage response proteins: optimizing protocols for super-resolution microscopy, *Methods* 48 (1) (2009) 63–71.
- [44] P. Constantinou, R.S. Dacosta, B.C. Wilson, Extending immunofluorescence detection limits in whole paraffin-embedded formalin fixed tissues using hyperspectral confocal fluorescence imaging, *J. Microsc.* 234 (2) (2009) 137–146.



- [45] T. Matsuzaki, H. Ozawa, Multiple immunofluorescence labeling in tissue sections, *J. Nippon Med. Sch.* 76 (6) (2009) 278–279.
- [46] R. Dawes, et al., Dax, a locust Hox gene related to fushi-tarazu but showing no pair-rule expression, *Development* 120 (6) (1994) 1561–1572.
- [47] R. Kelsh, I. Dawson, M. Akam, An analysis of abdominal-B expression in the locust *Schistocerca gregaria*, *Development* 117 (1) (1993) 293–305.
- [48] R. Kelsh, et al., Homeotic gene expression in the locust *Schistocerca*: an antibody that detects conserved epitopes in Ultrabithorax and abdominal-A proteins, *Dev. Genet.* 15 (1) (1994) 19–31.
- [49] F.F. Bier, et al., DNA microarrays, *Adv. Biochem. Eng. Biotechnol.* 109 (2008) 433–453.
- [50] S.L. Karsten, D.H. Geshwind, Gene expression analysis using cDNA microarrays, *Curr. Protoc. Neurosci.* Chapter 4 (2008) Unit4.28.
- [51] M. Stangegaard, Gene expression analysis using agilent DNA microarrays, *Methods Mol. Biol.* 529 (2009) 133–145.
- [52] R.B. Stoughton, Applications of DNA microarrays in biology, *Annu. Rev. Biochem.* 74 (2005) 53–82.
- [53] Z. Ma, et al., Modulation of behavioral phase changes of the migratory locust by the catecholamine metabolic pathway, *Proc. Natl. Acad. Sci. U S A* 108 (10) (2011) 3882–3887.
- [54] W. Guo, et al., Molecular cloning and temporal-spatial expression of I element in gregarious and solitary locusts, *J. Insect. Physiol.* 56 (8) (2010) 943–948.
- [55] F. Jiang, et al., Large-scale transcriptome analysis of retroelements in the migratory locust, *Locusta migratoria*, *PLoS One* 7 (7) (2012) e40532.
- [56] A.M. Ambrus, M.V. Frolov, The diverse roles of RNA helicases in RNAi, *Cell Cycle* 8 (21) (2009) 3500–3505.
- [57] M.E. Boisvert, M.J. Simard, RNAi pathway in *C. elegans*: the argonautes and collaborators, *Curr. Top. Microbiol. Immunol.* 320 (2008) 21–36.
- [58] D. Kim, J. Rossi, RNAi mechanisms and applications, *Biotechniques* 44 (5) (2008) 613–616.
- [59] J.B. Preall, E.J. Sontheimer, RNAi: RISC gets loaded, *Cell* 123 (4) (2005) 543–545.
- [60] S.E. Mohr, N. Perrimon, RNAi screening: new approaches, understandings, and organisms, *Wiley Interdiscip. Rev. RNA* 3 (2) (2012) 145–158.
- [61] G.B. Poulin, A guide to using RNAi and other nucleotide-based technologies, *Brief Funct. Genomics* 10 (4) (2011) 173–174.
- [62] E.H. Feinberg, C.P. Hunter, Transport of dsRNA into cells by the transmembrane protein SID-1, *Science* 301 (5639) (2003) 1545–1547.
- [63] W.M. Winston, C. Molodowitch, C.P. Hunter, Systemic RNAi in *C. elegans* requires the putative transmembrane protein SID-1, *Science* 295 (5564) (2002) 2456–2459.
- [64] Y. Dong, M. Friedrich, Nymphal RNAi: systemic RNAi-mediated gene knockdown in juvenile grasshopper, *BMC Biotechnol.* 5 (2005) 25.
- [65] J. Cabrero, et al., The Ku70 DNA-repair protein is involved in centromere function in a grasshopper species, *Chromosome Res* 21 (4) (2013) 393–406.
- [66] Y. Wang, et al., Second-generation sequencing supply an effective way to screen RNAi targets in large-scale for potential application in pest insect control, *PLoS One* 6 (4) (2010) e18644.
- [67] H. Huvenne, G. Smagghe, Mechanisms of dsRNA uptake in insects and potential of RNAi for pest control: a review, *J. Insect. Physiol.* 56 (3) (2010) 227–235.
- [68] H. Li, et al., Transcriptome analysis and screening for potential target genes for RNAi-mediated pest control of the beet armyworm, *PLoS One* 8 (6) (2013) e65931.