Published in final edited form as:

Dev Comp Immunol. 2014 July ; 45(1): 56–66. doi:10.1016/j.dci.2014.01.022.

No more non-model species: The promise of Next Generation Sequencing for Comparative Immunology

Nolwenn M Dheilly^{1,2,✉}, Adema Coen³, David A Raftos⁴, Gourbal Benjamin^{1,2}, Grunau Christoph^{1,2}, and Du Pasquier Louis⁵

¹CNRS, UMR 5244, Ecologie et Evolution des Interactions (2EI), Perpignan, F-66860, France

²Université de Perpignan Via Domitia, Perpignan, F-66860, France ³Center for Evolutionary and Theoretical Immunology, Biology Department, University of New Mexico, Albuquerque, NM 87131 USA

⁴Department of Biological Sciences, Macquarie University, North Ryde, NSW 2109, Australia

⁵University of Basel, Institute of Zoology and Evolutionary Biology, Basel, Switzerland

Abstract

Next generation sequencing (NGS) allows for the rapid, comprehensive and cost effective analysis of entire genomes and transcriptomes. NGS provides approaches for immune response gene discovery, profiling gene expression over the course of parasitosis, studying mechanisms of diversification of immune receptors and investigating the role of epigenetic mechanisms in regulating immune gene expression and/or diversification. NGS will allow meaningful comparisons to be made between organisms from different taxa in an effort to understand the selection of diverse strategies for host defence under different environmental pathogen pressures. At the same time, it will reveal the shared and unique components of the immunological toolkit and basic functional aspects that are essential for immune defence throughout the living world. In this review, we argue that NGS will revolutionize our understanding of immune responses throughout the animal kingdom because the depth of information it provides will circumvent the need to concentrate on a few “model” species.

1. Introduction

The vast majority of studies in immunology focus on medical or veterinary subjects, for obvious and justifiable reasons. The resulting paucity of data on immune responses in non-mammalian species has skewed our understanding of host defence in the vast majority of species on earth, leaving room for the erroneous interpretation that they are more “simple” than ourselves. However, work by comparative immunologists has revealed that the immune systems of non-mammalian species (particularly invertebrate animals) are not only much more complex than previously assumed but can also vary much more among classes or phyla (Loker et al. 2004). This fits with evidence from whole genome sequencing studies, which have shown that the number of expressed genes per genome is roughly equivalent in most multicellular animals. The obvious conclusion is that the genomic playing field

✉ Corresponding authors: Dheilly Nolwenn Marie, Université de Perpignan via Domitia, UMR 5244 CNRS, Ecologie et Evolution des Interactions, 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France, +33 (0) 6 25253514, ndheilly@hotmail.fr.

available for the development of complex immune systems in different taxa is much more level than previously assumed.

Despite advances at the level of genomics, accompanying information about the physiological function of immune response genes in non-model species is often lacking. In most cases, we still do not understand the biological relevance of the gene systems that appear to be associated with host defense, how they help organisms to combat infection, how they evolved, or even whether they support immune responses to infection that are vaguely comparable to our own. To date, partial answers to these questions have come from other experimental approaches. Ecological studies have investigated interaction between hosts and their symbionts (ranging from parasites to mutualists) at the level of whole organisms or populations.

One of the main challenges now for evolutionary immunologists is to link the molecular systems that they have detected using genomics (and transcriptomics) with these newly identified forms of immune response. There are three outstanding questions: (1) What is the fundamental genetic toolkit of the immune system? In other words, which are the core genes and gene networks that underpin immune responses throughout the animal kingdom? And by corollary, (2) Which immune response genes evolved *de novo* in individual taxa, and what do they do? For ecological immunologists, the main goal is (3) to link particular ecological interactions with the selection of immune response pathways (Schulenburg et al. 2009). That is, what are the core ecological interactions (host-parasite, host-symbiont) that lead to the genome structure of immune response genes?

We believe that the answers to these fundamental questions about the evolution of immune systems will come from a combination of data from next generation nucleotide sequencing (NGS) and experimental ecology. To support that argument, this article discusses our current understanding of invertebrate immune systems and then describes how different applications of NGS can be used to further that understanding. We place special emphasis on invertebrate immune responses because our knowledge of these organisms has been limited by a lack of genome and transcriptome data when compared to mammals.

Invertebrates represent roughly 90% of the planet's animal species. The basic understanding of invertebrate immune responses will have significant impacts for many of the major societal challenges that we face today. This has already been the case when one considers for instance that a recent attribution of Nobel prizes in physiology and medicine acknowledges the importance of comparative immunology for medical sciences (Imler and Ferrandon 2011). There are also new challenges, such as global warming. In the near future, we can expect global environmental changes to afford new environments that will favour the apparition, or spread of numerous diseases that will impact all marine and terrestrial ecosystems (and a great majority of invertebrate species) (Patz et al. 2000; Harvell et al. 2002; Lafferty 2009). These changes will affect human health and society (Campbell-Lendrum et al. 2007): vector borne diseases transmission, agriculture (biological pests controls, and pests), and aquaculture will be impaired to some unpredictable extent. More than ever, comparative approaches will be welcome.

1.1. Our current understanding of invertebrate immune systems

The study of invertebrate immune systems has a broad and rich history. It is almost obligatory to point out that the first experiments in modern immunology were conducted by Metchnikoff, not in humans or mice, but in sea star larvae, waterfleas and newts and that his work led to the discovery of encapsulation and phagocytosis (Metchnikoff 1893). In the intervening 120 years, the field of comparative immunology made substantial advances by investigating invertebrate immune responses at a cellular level (including studies of phagocytosis, cytotoxicity and the production of antimicrobial peptides) and by tracing the evolutionary history of specific immune response gene families, such as complement component C3 and its related thioester containing proteins (TEP). (Loker, Adema et al. 2004)

However, until very recently, comparative immunology has been hindered by the lack of broad sequence data that can be used to make comparisons across the level of entire genomes, rather than just specific genes (Adams et al. 2000). In the past five years, our most significant breakthroughs have been the direct consequence of studying the diversity of invertebrate immune responses at the genomic level (Rast and Messier-Solek 2008; Messier-Solek et al. 2010). At first, and in accordance with the limited number of sequenced genomes, our view of animal immunity and its evolution came almost entirely from investigations in mammals birds, amphibians fish and insects (especially *Drosophila melanogaster*). This selective representation of deuterostome vertebrates and ecdysozoan protostome invertebrates, to the exclusion of prebilaterian and lophotrochozoan protostomes and invertebrate deuterostomes together with incomplete physiological approaches to immunity, gave the impression that vertebrates had highly complex immune systems combining innate and adaptive immunity, whilst invertebrates were “poor cousins” with simple innate immune system.

The wave of new nucleotide sequences that have become available in the past five years (Figure 1) has provided a very different and complex picture of the evolution of immune systems throughout the animal kingdom (Rast and Messier-Solek 2008). Briefly, the most up-to-date evidence suggests that early metazoans such as sponges and cnidarians (Putnam et al. 2007; Chapman et al. 2010; Shinzato et al. 2011) already possessed all the gene families usable by immune systems, signifying a common eumetazoan inheritance for the basic immune response toolkit (Hemmrich et al. 2007; Miller et al. 2007; Kvennefors et al. 2008; Srivastava et al. 2010). In this context, it appears that the less elaborate innate immune systems evident in the genomes of some urochordate deuterostomes (*Ciona intestinalis*, *Oikopleura dioica*), certain ecdysozoans (such as *Caenorhabditis elegans*), and some insect species, are the result of gene loss within these lineages (Kortschak et al. 2003; Denoeud et al. 2010). When the broader phylogenetic picture revealed by genome sequencing is taken into account, a remarkable expansion of innate immune receptors and effectors is evident in basal deuterostomes, such as sea urchins (echinoderms), and basal chordates, such as amphioxus (cephalochordates) and a number of protostome taxa (Hibino et al. 2006; Rast et al. 2006; Huang et al. 2008). For instance, a significant expansion of Toll-like receptor diversity has been found in annelid worms (protostome invertebrates), suggesting that

selection pressures to massively expand receptor repertoires have been present throughout the protostome and deuterostome lineages of animal evolution (Davidson et al. 2008).

This trend toward diversification is evident among not just receptors associated with “innate” immune systems, but also in hypervariable recognition molecules generated somatically akin to the antibodies and T-cell receptors of vertebrates. Over the past 10 or so years, highly variable molecules that seem to participate in new forms of inducible, pathogen specific immune responses have been found in a broad range of invertebrate taxa, including sea urchins, molluscs, insects and crustaceans even though their role is far from being fully understood. These molecules include 185/333 proteins from sea urchins (Nair et al. 2005; Terwilliger et al. 2007), FREPs from molluscs (Adema et al. 1997; Zhang and Loker 2004; Zhang et al. 2008), penaeidins in crustaceans (Cuthbertson et al. 2002), and DSCAMs in insects and crustaceans (Watson et al. 2005; Schmucker and Chen 2009; Dong et al. 2012). Their discovery raise fundamental questions such as where (and when from an evolutionary point of view) does innate immunity stop and adaptive immunity begins? (Ziauddin and Schneider 2012). All of these systems are now being thoroughly investigated to understand their evolutionary origins, the ecological circumstances of their diversification and their functional links with other immune regulatory pathways.

At the same time that molecular genetic technologies were revealing the existence of highly variable immune response genes in invertebrates, studies of host-parasite and host-symbiont co-evolution by ecological immunologists began to reveal unexpected results at the level of whole organisms and populations. These data have allowed us to propose a model that predicts how host-pathogen interactions shape the immune system. This model suggests that symbionts (parasites, commensalists or mutualists) all exert some negative effects on the physiology of their host (Figure 2). In some cases, they also provide beneficial effects that may compensate for their negative impact on the host fitness. Confronted to this situation the host offers a proportional evolutionary response to counter infection, or limit its pathogenicity. Within this gamut, active immune responses to prevent infection are seen as expensive physiologically. They place demands on the energy budget of cells and individuals, reducing the availability of energy that can be allocated to reproduction, growth and other key functions (Sheldon and Verhulst 1996; Stahlschmidt et al. 2013). Hence, from an ecological perspective, hosts are most likely to evolve immune responses that trade-off regulation of their populations of symbionts (from parasites to mutualists) with conserving an optimum fitness.

Such a trade-off can sustain both the elimination of pathogens and the tolerance of symbionts (Medzhitov et al. 2012). The arms race of host response and pathogen virulence that plays out during host-pathogen co-evolution drives the diversity of gene families and gene polymorphism within populations (Eizaguirre et al. 2012; Kubinak et al. 2012; Mitta et al. 2012). Host-mutualist co-evolution has led to a distinct phenomenon, specialisation of immune responses (Reynolds and Rolff 2008; Login et al. 2011). Thus, the intensity and type of host response result from trade-offs that are required to assure host fitness during their co-evolution with pathogens, or other commensals (Figure 2). However, the direct demonstration that host symbiont coevolution lead to either diversification of specialization of the immune system are very scarce and the proposed model on host parasite evolution is

based on only few model species. More comprehensive sampling is now necessary to confirm that this model point to generalities.

1.2. Next generation sequencing and its role in addressing the key questions facing evolutionary immunology

We have shown that even though our current understanding of invertebrate immune systems at the molecular level is strongly biased towards few model species, genome sequencing has proven its power in revealing the diversity of immune system genome structures. Ecological immunology has provided some evidence that ecological factors have a key role in the evolution of immunity, which result in very diverse immune systems (Schulenburg, Kurtz et al. 2009). However, in most cases, the molecular mechanisms responsible for the observed phenotypes have not been characterized due to limited access to molecular biology approaches for non-model species. Next generation sequencing is now emerging as an essential tool for such cross-disciplinary research. The comprehensive sampling, with indication of abundance levels, at modest per read effort renders NGS superior to Sanger sequencing because it facilitates integrative approaches and broad, large scale comparisons among different organisms (Table 1). It can be applied to population biology, ecology, evolutionary biology, and molecular biology. NGS is now dominated by four main techniques developed by different companies: Roche Applied Science (454 Genome Sequencer FLX System, Branford, CT, USA), Illumina (Genome analyser, San Diego, CA, USA), Helicos BioSciences (HeliScop Single Molecule Sequencer, MA, USA) and Life technologies (Sequencing by Oligonucleotide Ligation and Detection; SOLID, Carlsbad, CA, USA). The technical details of these different systems have already been extensively reviewed elsewhere and so they will not be described further here (Hudson 2008; Morozova and Marra 2008; Shendure and Ji 2008). The length of individual sequences (reads) produced by NGS is shorter than Sanger sequencing, but millions of reads can be generated in a short period of time, providing massively increased depth of sequencing (Table 1). NGS can generate three to four order of magnitude more sequences than traditional methods and is considerably less expensive (Table 1). As a result, NGS will benefit numerous fields of research, especially with the continued improvement of computational resources to manage and analyse the large datasets collected. We believe that the contribution of NGS will be especially important in our understanding of evolutionary immunology (Figure 3).

2. Transcriptome sequencing

2.1. De novo transcriptome assembly

The most straightforward contribution that NGS will make to evolutionary biology is to provide either genomic or transcriptomic datasets for a very broad diversity of species across all phyla. Even though NGS has now replaced standard Sanger sequencing as gold standard for whole genomic DNA sequencing, (Li et al. 2010), this application remains relatively expensive and requires considerable expertise, as well as complex infrastructure for data collection and analysis. As an alternative, *de novo* assembly of NGS transcriptome data significantly reduces the size of the target space because it relies on sequencing cDNA (representing transcribed genes) rather than genomic DNA. Initially, NGS transcriptomes had to be assembled *de novo* using the genome of a closely related species as a reference or

“scaffold” (e.g. the wasp *Polistes metricus* (Toth et al. 2007)). However, continued computational developments have followed in rapid order to improve both the length of NGS sequence reads and the capability of assembler software. This has led to several software packages that can assemble transcriptomes from short sequence reads without a reference genome, allowing deep transcriptomic analyses of numerous species that do not have closely related reference genomes. The first of these *de novo* assemblies was performed in transcriptome data for various life history stages of the Glanville fritillary butterfly (*Melitaea cinxia*) (Vera et al. 2008).

In many invertebrate species, multigenic families with a direct or indirect role in immunity present a greater number of loci than the same families in vertebrates, which increases the difficulty in reliably assemble *de novo* the transcripts or genes. Effective *de novo* assembly is now also feasible for short NGS reads with high sequence similarity that may encode transcripts from multiple alleles of the same gene, different members of the same multilocus family, or highly variable immune response genes generated by post-genomic processes. For example, the diversity of fibrinogen-related proteins (FREPs) in several strains of *B. glabrata* had previously been assessed by extensive Sanger sequencing of cloned PCR amplicons and BAC inserts (Zhang and Loker 2003; Zhang and Loker 2004). In a single whole transcriptome NGS experiment (Illumina sequencing), bioinformatics analysis led to doubling the number of FREP gene subfamilies that had been detected previously by traditional Sanger sequencing (Dheilly et al. Submitted). Moreover, the high levels of sequence diversity among the FREP sequence fragments that were detected by NGS from an additional strain of *B. glabrata* suggested that yet more gene families remain to be discovered.

These results are highly significant for comparative immunology because they confirm that the levels of diversity within immune response gene families in invertebrates may approach those of comparable gene systems in mammals. In addition, the data demonstrate that *de novo* NGS transcriptome sequencing can detect and help us to identify novel, highly diversified immune response gene families in other taxa from which complete genome sequence are lacking. For example, a great diversity of antimicrobial peptides was recently discovered in the ladybeetle *Harmonia axyridis* (Vilcinskas et al. 2013). Now, with NGS, such descriptive approaches can be coupled to studies aiming at finding out the role of this diversity at the biological level. However, we wish to remind how important it is to couple genome analysis with functional analysis. For instance, gene from the same family may participate in distinct physiological functions, such as Tolls and Dscams that may be involved in immune response but also in neurones development and central nervous system functions (Schmucker et al. 2000; Dong et al. 2006; McIlroy et al. 2013).

2.2. Dual *de novo* assemblies

‘Dual-*de novo*’ assemblies, in which both parasite and host transcriptome are assembled from the same sample, may also be useful to provide a more comprehensive understanding of the immune processes that play out during host-pathogen interactions.

Dual assemblies will allow inducible immune response genes to be identified by comparing the transcriptomes of host organisms collected from naturally-infected populations, or from

host experimentally exposed to pathogens with the resting transcriptome of unchallenged hosts that include “only” constitutively-expressed innate response factors. For species that host parasites with complex life-cycles, the dual *de novo* assembly approach will allow the transcriptomes for different stages of parasite life-cycle to be compared. From a more integrated point of view, a comprehensive understanding of immune gene expression at a given time would in theory necessitate characterizing all symbionts (microbial community, viruses and eucaryote parasites) in the individual under study. It is often difficult to diagnose latent pathogens in many invertebrates because most pathogens can neither be cultured, purified, nor be found outside their host organisms. Hence, it is anticipated that many novel pathogens or symbionts will be identified as “stowaway passengers” while sequencing the transcriptomes or genomes of their hosts. This approach has already been used to identify a novel totivirus associated with cardiomyopathy syndrome in salmon (Lovoll et al. 2010) and is currently being used to search for unidentified pathogens associated with mortalities in pearl oysters (DA Raftos, personal communications).

Despite the success of this approach, it is acknowledged that the presence of transcripts from multiple species of distant phylogenetic identity within a single sample may negatively impact the quality of the *de novo* transcriptome assembly. It may also be difficult to reliably assign a transcript to either the host or symbiont (Hraber and Weller 2001; DeJong et al. 2004). Taxonomic assignment may be performed based on (i) tblastx e-value, (ii) the percentage of GC base content and (iii) species specific signatures based on CLaMS (Sabourault et al. 2009; Pati et al. 2011; Zhuang et al. 2012; Vidal-Dupiol et al. 2013). Moreover, the routine purification of transcripts for analysis of eukaryote transcriptomes frequently relies on capture of polyA tails, whereas the study of prokaryotes relies on purification of micro-organisms followed with sequencing of total RNA. Hence, it remains to be demonstrated whether multiple *de novo* assembly of host and microbiota can be performed routinely on samples from eucaryote hosts containing prokaryotic pathogens, since it will necessitate enrichment of different RNA species (Westermann et al. 2012).

3. Association studies

3.1. Whole genome association studies

One of the primary reasons for the use of NGS is the re-sequencing of whole genomes to identify variations or mutations associated with a particular phenotypic trait. However, such approach remain costly and limited to model species for which the genome has been sequenced. At lower cost, whole transcriptome sequencing also allows the identification of single nucleotide polymorphism (SNP) and estimation of allele frequency in genes with sufficiently high expression level (when performed in the appropriate tissue for the trait of interest) (Cirulli et al. 2010). Numerous software tools are now available for the detection of SNPs in NGS data sets (Marth et al. 1999; Li et al. 2008; Quinlan et al. 2008; Koboldt et al. 2009; Li et al. 2009; Shen et al. 2010). The association between specific SNPs and specific immune traits means that comparative immunologists need to pay particular attention to SNPs within host populations. Such phenotype-association studies have yielded a broader understanding of the genetic complexity in responses to a range of immune insults, including vaccination (Biscarini et al. 2010; Pankratz et al. 2010), Lipopolysaccharides

(Biscarini, Bovenhuis et al. 2010), herpesvirus (Kongchum et al. 2010), cancer (Ho-Pun-Cheung et al. 2010), meningitis (Da Silva et al. 2011), and mastitis in cattle (Carvajal et al. 2013). However, among invertebrates immune-related SNPs have been localised but phenotype-association studies have been rarely undertaken (Cohuet et al. 2008; Nunez-Acuna and Gallardo-Escarate 2013). The use of NGS will provide a more convenient and cost-effective method for discovery of SNP in non-model species and so could provide important information about the genetic basis of inter-individual variation in immune responses. More specifically, population genomic approaches that incorporate SNP analysis can identify genetic markers for disease resistance, as demonstrated by the NGS approach taken by Bangham et al. (Bangham et al. 2007). The authors identified a mutation associated with resistance to sigma virus in *Drosophila* (Bangham, Obbard et al. 2007). Such approaches may be extremely valuable for selective breeding programs in industries. In addition, transcriptome analysis allows identifying Single nucleotide variants (SNV) within individuals and has recently revealed extensive RNA editing in humans (Chepelev et al. 2009; Peng et al. 2012; Lee et al. 2013). Again phenotype-association studies has been used to reveal its role in cancers (Chepelev, Wei et al. 2009).

3.2. Targeted sequencing

NGS may also be used in a selective fashion to sequence a limited number of loci, yielding a high read coverage of the sequencing targets. This powerful approach has been employed to characterize the olfactory receptor gene family, the largest multigene family in mammals (Hughes et al. 2013). It can also be used for population level genetic analyses. Primer tags facilitate barcoding transcripts from different samples for multiplex sequencing in order to genotype loci of interest in a large number of individuals (Binladen et al. 2007). This approach has been used to characterise the highly polymorphic major histocompatibility complex (MHC) of the bank vole, *Myodes glareolus* (Babik et al. 2009). It allowed alleles that were present at very low frequency within the population to be detected and identified a significant association between certain MHC alleles and the intensity of infection by the pinworm parasite, *Aspiculuris tetraptera*.

Similar approaches may be particularly useful among invertebrates. Recent evidence suggests that invertebrates encode numerous categories of immune-related factors that appear to use high levels of sequence diversity to recognize the antigenic proteins or Microbial-associated molecular patterns (MAMPs) of parasites (Ausubel 2005). This diversity often appears to be associated with polymorphic, multigenic families that are further diversified by post-genomic mechanisms such as somatic mutations (including recombination, gene conversion), alternative splicing and RNA editing. NGS is ideal for studying such systems whether at the RNA or DNA levels. For instance, targeted NGS sequencing of the highly diverse CDR3 (complement determining region 3) of T-cell receptors in human peripheral blood leukocytes increased by an order of magnitude the number of known variants from 3187 to 33,664 TCR β mRNA sequences (Freeman et al. 2009). Comparable approaches could be used to investigate the diversity of highly variable immune-related proteins of invertebrates. In particular, NGS could test whether the repertoire of antigen receptors in invertebrates is static, or whether it is continually modified over time or in response to immune challenge. Several candidate systems, which already

have detailed information from one species and indications of the presence of homologous genes in close relatives, are already available for such analyses. For instance, *Anopheles gambia* responds to infection by producing antigen-specific splice forms of Dscams (Dong, Cirimotich et al. 2012). In the purple sea urchin, *Strongylocentrotus purpuratus*, different repertoires of 185/333 proteins are expressed in response to different challenges (Terwilliger, Buckley et al. 2007; Dheilly et al. 2009), whilst in the pond snail, *B. glabrata*, there is evidence that sequences from different FREP families are expressed in response to different forms of immune challenge (Zhang, Yong et al. 2008; Adema et al. 2010). In these systems, the use of targeted NGS to identify SNP and SNV to profile the diversity of receptor repertoires over time, in different tissues, and in response to immune challenge will provide information about the levels of specificity in these defence responses and the mechanisms by which those repertoires change (alternative splicing, RNA editing, etc.).

4. Gene expression profiling

4.1. Comparative transcriptomic studies

Perhaps the biggest contribution that NGS will make to our understanding of invertebrate immune responses is its capacity to identify genes from across entire genomes that are either up- or down-regulated during immune responses to particular types of infectious agent. For the first time, NGS affords textured, genome-wide assessment of the intricacies of complex immune responses at the level of the transcriptome. One shortfall of genome sequencing and subsequent gene annotation in invertebrates is that substantial proportions of the putatively expressible genes are novel or are not sufficiently conserved so that orthologues can be easily identified in sequence databases. Accordingly, no function can be assigned based on similarity with previously studied genes. In these cases, the increased expression of a novel transcript in response to immune challenge detected by NGS-based high throughput transcriptomic profiling can be used to provide an initial indication that the encoded protein has an immune responsive function. Such transcriptomic profiling can also be used to provide information on other physiological functions that are impacted over the course of an immune challenge (including the messages that are down regulated). Previously, microarray analyses were performed extensively to provide these types of transcriptional profiles (De Gregorio et al. 2001; Aujame et al. 2002; Hutton et al. 2004; Schweitzer et al. 2010) and identify candidate sequences of putative immune factors for downstream functional studies. In those downstream studies, expression profiles of specific target genes have most often been assessed over the course of immune responses using quantitative RT-PCR (qRT-PCR). However, this approach has been limited by the relatively shallow depth of sequence diversity that can be assessed using cDNA microarrays. In contrast, mRNA sequencing by NGS generates millions of cDNA fragments (reads) and mRNA expression can be rapidly evaluated by counting the number of reads that match the gene or transcript of interest in the corresponding genome or transcriptome, even though candidate sequences still require functional validation of immune function and involvement (Galinier et al. 2013). The ever decreasing cost of RNA sequencing also implies that kinetic studies are being possible, which will considerably enhance biological significance of gene expression studies.

In parallel to the dual-*de novo* assembly of transcriptome approach discussed above, it will also be useful to adopt 'dual RNA-seq' approaches, in which the expression profiles of both pathogens and their hosts can be assessed simultaneously (Westermann, Gorski et al. 2012). The complexity of such studies will depend on the nature of the host/pathogen interaction, due to the need for different RNA enrichment protocols for different types of pathogens (eukaryote, prokaryote or virus), and the sequencing depth needed for accurate coverage of both host and pathogen (Westermann, Gorski et al. 2012). For example, Juranic Lisnic et al. (Juranic Lisnic et al. 2013) recently demonstrated the utility of this approach by studying virus-host cell interactions of murine Cytomegalovirus.

4.2. micro RNA

Noncoding RNAs (small nuclear or nucleolar RNAs and microRNAs), which produce functional RNA molecules rather than encoding proteins are involved in regulating a broad range of physiological processes, including immune responses (Eddy 2001; Carpenter et al. 2013; Curtale and Citarella 2013; Minton 2013). It has recently been demonstrated that small non-coding RNAs are responsible for post-transcriptional regulation of gene expression either via the degradation of target mRNAs or the inhibition of protein translation. For instance, exposure of human cell lines to LPS has been shown to induce microRNAs (Taganov et al. 2006) that specifically regulate the NF- κ B signalling pathway (Taganov, Boldin et al. 2006). More recently, a microRNA (miR-29) was found to specifically target IFN- γ in humans and to suppress immune response to intracellular pathogens (Ma et al. 2011). Antiviral immunity in bacteria, plants and invertebrates also involves the production of small RNAs that interfere with viral replication (Ding and Voinnet 2007; Sorek et al. 2008). The ability of host microRNAs to evolve rapidly (Meunier et al. 2013) makes them ideal candidates for the control of host-pathogens interactions. However, pathogens, including viruses also produce microRNA that can help them evade the immune system (Sullivan 2008). Hence there seems to be a molecular arms race between virus and host genomes involving microRNAs (Ding and Voinnet 2007).

NGS strategies already exist to capture and analyse the complement of microRNA sequences in non-model species (Buermans et al. 2010). The identification of microRNA sequences relies on appropriate data analysis. Since all sequence reads produced from any given NGS platform are now longer than the average microRNA, it initially relies on finding the 3' and 5' ends of the microRNA in the same read and then to perform secondary structure analysis. Using this approach, RNA-seq successfully identified novel microRNAs in tomatoes (Moxon et al. 2008), human embryonic stem cells (Bar et al. 2008), chicken embryos (Glazov et al. 2008) and mononuclear cells from peripheral blood (Vaz et al. 2010). These NGS studies confirmed the identities of almost all previously known microRNA, and identified hundreds of new candidate microRNAs. In the context of host-pathogen interactions, RNAseq has also been employed to study microRNAs in HeLa cells responding to challenge by *Salmonella* (Schulte et al. 2011).

4.3. Alternative splicing

Alternative splicing of mRNAs is one of the best known mechanisms responsible for the generation of post-genomic diversity (Nadal-Ginard et al. 1991). However, we are only

starting to understand the regulatory processes that govern splicing decisions and how this regulation shapes the splicing phenotypes observed in different tissue types and developmental stages. Cell surface receptors involved in the nervous system and in immune responses are among the genes that are most frequently associated with alternative splicing (Modrek et al. 2001). In plants, alternative splicing plays a key role in providing transcriptome plasticity necessary to better cope with stress and pathogens (Mastrangelo et al. 2012). In mammals, alternative splicing plays an extensive role in regulation of T-cell activation (Lynch 2004; Ip et al. 2007), whilst in insects Dscams exhibit pathogen-specific splice-form expression following infection with different pathogens (Dong, Taylor et al. 2006). Both insects and crustaceans use alternative splicing to produce tens of thousands of different Dscams sequence variants (Watson, Puttmann-Holgado et al. 2005; Brites et al. 2008). Similarly, in the pond snail, *B. glabrata*, alternatively spliced FREPs have been identified (Zhang and Loker 2003), yet their functional significance remains to be analysed.

The systematic analysis of alternative splicing used to be undertaken using expressed sequence tags (EST) or specialized microarrays. However, EST are subject to cloning biases and generally provide only low coverage, and the specificity of microarrays is affected by cross-hybridization. NGS RNA-seq data can now be used as a more effective alternative to identify novel splice junctions. NGS has been employed previously to identify new alternative splicing sites in humans (Pan et al. 2008; Wang et al. 2008), *C. elegans* (Ramani et al. 2010) and *Plasmodium falciparum* (Sorber et al. 2011). Specialized tools for splice junction identification are available but most necessitate a reference genome (Trapnell et al. 2009). Only a few software packages, such as SplitSeek, are able to identify splice sites in uncharacterized transcripts (Ameur et al. 2010). In the future, the development of additional software that allow more automated identification of splicing sites from the *de novo* assembled NGS transcriptomes will have significant outcomes.

Other approaches, such as cross-linking immunoprecipitation sequencing (CLIP-Seq) and RNA immunoprecipitation sequencing (RIP-Seq) have been used to map RNA-binding sites for splicing factors (Sanford et al. 2009; Yeo et al. 2009; Zisoulis et al. 2010). These techniques have great potential to examine mRNA processing in a large range of invertebrate species that generate highly variable molecules over the course of an immune response (see above).

4.4. Transcription factors

Chromatin immunoprecipitation coupled with NGS, known as ChIP-Seq, has been employed to identify protein-DNA interactions. In particular, it has been extensively used to identify the DNA binding sites of transcription factors. Transcription factors are required for activation of immune cells, and so it is necessary to identify their targets to comprehensively understand how they influence gene transcription and cell fate. For example, ChIP-Seq has been used to identify the binding sites of the NF κ B factor p65 on human chromosome 22 (Martone et al. 2003) and the targets of STAT4 and STAT6 transcription factors involved in T helper cell differentiation (Wei et al. 2010).

5. Epigenetic profiling

Epigenetic modifications are mitotically or meiotically heritable changes in gene expression that are not based on alterations of nucleotide sequence and are potentially reversible (metastable). Such modifications can impact gene expression, alternative splicing and other mechanisms of diversification in immune response genes. To date, three mechanisms of epigenetic change are known to play key roles in the regulation of immune responses: Protein-nucleic acid interactions, histone modifications and DNA methylation.

5.1. Histone modifications

NGS techniques related to CHIPseq such as native-ChIP (N-ChIP) and crosslink-ChIP (X-ChIP) capture DNA-histone interactions (O'Neill and Turner 2003; Cosseau et al. 2009). The distribution of histone modifications identified using these methods can now be found in several databases e.g. (Zhang et al. 2010). Histone modifications are employed by both host and pathogens to regulate gene expression. For instance, the bacterial pathogen, *Listeria monocytogenes*, has been shown to induce dephosphorylation of histone H3 and a deacetylation of histone H4 during infection in mammals, thus reducing the transcriptional activity of some key immune genes (Hamon et al. 2007). Similarly, it has been shown that histone H3 phosphorylation by IKK- α is a key step in the NF κ B pathway of immune activation (Yamamoto et al. 2003). ChIP-seq has also revealed that histone modifications are involved in antigen variations among pathogens such as *Schistosoma mansoni* (Perrin et al. 2013).

5.2. DNA methylation

In mammals, DNA methylation is associated with the control of gene expression and the maintenance of cell lineages. It is necessary for T cell activation, proliferation, memory cell formation and activation. DNA methylation appears rapidly post infection and is maintained thereafter (Weng et al. 2012; Kondilis-Mangum and Wade 2013). They constitute stable markers of immunostimulation that facilitate more rapid cellular responses upon reinfection. The role of DNA methylation in mammalian B cell lineages formation is also being investigated and may be linked to activation-induced deaminase (Kondilis-Mangum and Wade 2013). In addition, substantial DNA methylation has been detected in many invertebrate species ranging from deuterostomes (*S. purpuratus* (Bird et al. 1979)) to molluscs (*C. gigas* (Gavery and Roberts 2010) and *B. glabrata* (Fneich et al. 2013)), insects (*A. mellifera* (Lyko et al. 2010)), and tunicates (*C. intestinalis* (Suzuki et al. 2007)). In contrast, DNA methylation appears to be more limited in other invertebrates, such as *D. melanogaster* (Gowher et al. 2000) and *C. elegans* (Simpson et al. 1986). Although the data are limited, DNA methylation may also play important roles in the development and phenotypic plasticity of invertebrates (Regev et al. 1998; Roberts and Gavery 2012). However, to date, the role of DNA methylation in the regulation of invertebrate immune response genes has not been investigated.

Three NGS techniques have been developed in order to study the DNA methylome. In BS-Seq, bisulphite conversion of the DNA is followed by whole genome sequencing (Laurent et al. 2010). In MeDIP-Seq, DNA is fragmented and regions containing methylated DNA are

immunoprecipitated with 5-methylcytosine antibodies before NGS (Down et al. 2008), whilst in MBD-Seq, DNA fragments containing CpG's specifically interact with methyl-binding domain conjugated on beads and are sequenced with NGS (Serre et al. 2010). The last two techniques are cost-effective methods to identify methylation-enriched genomic regions. For example, MeDIP-Seq has very recently been employed to identify aberrantly methylated genes in skin lesions from human patients with *Psoriasis vulgaris* infections (Zhang et al. 2013).

The ratio of observed to expected CpG dinucleotides has also been employed to predict methylation status in genomes and transcriptomes from many taxonomic groups (Shimizu et al. 1997; Elango and Yi 2008; Elango et al. 2009; Gavery and Roberts 2010). In invertebrates, CpG methylations are preferentially located within coding regions (Suzuki, Kerr et al. 2007). Therefore, the ratio of CpG dinucleotides can be calculated from transcriptomes. Recently, this approach was used to predict the methylation status of *B. glabrata* from a transcriptome generated *de novo* from RNAseq data (Fneich, Dheilly et al. 2013). This suggests that generating *de novo* a high number of transcriptomes from various species will also significantly enhance our understanding of the evolution of DNA methylation.

6. Conclusions and perspectives

The power and comparatively low cost of NGS is revolutionizing genomic and transcriptomic research. It provides opportunities to undertake comprehensive analyses of the biology of any organism not just a limited number of “model” species. For example, comparison of RNA-seq data from closely related species with different life history traits, lifespan, diet, reproductive strategies, symbionts and pathogens will provide significant information to link phenotypic traits such as the diversity of the immune system to evolution. The frequently documented rapid evolution of immune-related genes and the identification of novel highly variable immune response gene families from invertebrates (Ghosh et al. 2011) have challenged the technical capabilities of traditional investigative research methods in ways that limited our perspectives on the evolution of immune systems. The advent of NGS now provides for a new generation of comprehensive and integrative studies in comparative immunology. These encompass studies ranging from the investigation of immune regulation via epigenetic mechanisms, through gene discovery, to studies of gene expression and the diversification of immune factors. We also wish to emphasize that every time a immunome of a new species is characterized, someone drops on something special that could have medical implications. It is still daunting to consider the effort that will be needed to collect and analyze NGS datasets from numerous species to provide a broad representing of metazoan taxa. This may require concerted effort through the establishment of consortia of researchers interested in comparative immunology across a broad range of phyla. However, the technical feasibility afforded by NGS holds great promise and opportunity. In the future of comparative immunology there will be no more non-model species.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

CMA acknowledges support from NIH grant number P20GM103452 from the National Institute of General Medical Sciences (NIGMS). NMD was supported by the Agence Nationale de la Recherche Blanc, SVSE7, project Bodyguard.

References

- Adams MD, Celniker SE, et al. The Genome Sequence of *Drosophila melanogaster*. Science. 2000; 287(5461):2185–2195. [PubMed: 10731132]
- Adema CM, Hanington PC, et al. Differential transcriptomic responses of *Biomphalaria glabrata* (Gastropoda, Mollusca) to bacteria and metazoan parasites, *Schistosoma mansoni* and *Echinostoma paraense* (Digenea, Platyhelminthes). Mol Immunol. 2010; 47(4):849–860. [PubMed: 19962194]
- Adema CM, Hertel LA, et al. A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. Proc Nat Acad Sci. 1997; 94(16):8691–8696. [PubMed: 9238039]
- Ameur A, Wetterbom A, et al. Global and unbiased detection of splice junctions from RNA-seq data. Genome Biol. 2010; 11(3):R34. [PubMed: 20236510]
- Aujame LUC, Burdin N, et al. How microarrays can improve our understanding of immune responses and vaccine development. Ann NY Acad Sci. 2002; 975(1):1–23. [PubMed: 12538150]
- Ausubel FM. Are innate immune signaling pathways in plants and animals conserved? Nat Immunol. 2005; 6(10):973–979. [PubMed: 16177805]
- Babik W, Taberlet P, et al. New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. Mol Ecol Resour. 2009; 9(3):713–719. [PubMed: 21564729]
- Bangham J, Obbard DJ, et al. The age and evolution of an antiviral resistance mutation in *Drosophila melanogaster*. Proceedings of the Royal Society B: Biological Sciences. 2007; 274(1621):2027–2034.
- Bar M, Wyman SK, et al. MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. Stem Cells. 2008; 26(10):2496–2505. [PubMed: 18583537]
- Binladen J, Gilbert MTP, et al. The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. PLoS ONE. 2007; 2(2):e197. [PubMed: 17299583]
- Bird AP, Taggart MH, et al. Methylated and unmethylated DNA compartments in the sea urchin genome. Cell. 1979; 17(4):889–901. [PubMed: 487434]
- Biscarini F, Bovenhuis H, et al. Across-line SNP association study of innate and adaptive immune response in laying hens. Animal Genetics. 2010; 41(1):26–38. [PubMed: 19781038]
- Brites D, McTaggart S, et al. The Dscam homologue of the crustacean *Daphnia* is diversified by alternative splicing like in insects. Mol Biol Evol. 2008; 25(7):1429–1439. [PubMed: 18403399]
- Buermans H, Ariyurek Y, et al. New methods for next generation sequencing based microRNA expression profiling. BMC Genomics. 2010; 11(1):716. [PubMed: 21171994]
- Campbell-Lendrum D, Corvalan C, et al. Global climate change: implications for international public health policy. Bull World Health Organ. 2007; 85(3):161–244.
- Carpenter S, Aiello D, et al. A long noncoding RNA mediates both activation and repression of immune response genes. Science. 2013; 341(6147):789–792. [PubMed: 23907535]
- Carvajal A, Huircan P, et al. Single nucleotide polymorphisms in immunity-related genes and their association with mastitis in Chilean dairy cattle. Genet Mol Res. 2013; 12(3):2702–2711. [PubMed: 23979895]
- Chapman JA, Kirkness EF, et al. The dynamic genome of *Hydra*. Nature. 2010; 464(7288):592–596. [PubMed: 20228792]

- Chepelev I, Wei G, et al. Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. *Nuc Acids Res.* 2009; 37(16):e106.
- Cirulli E, Singh A, et al. Screening the human exome: a comparison of whole genome and whole transcriptome sequencing. *Genome Biol.* 2010; 11(5):R57. [PubMed: 20598109]
- Cohuet A, Krishnakumar S, et al. SNP discovery and molecular evolution in *Anopheles gambiae*, with special emphasis on innate immune system. *BMC Genomics.* 2008; 9(1):227. [PubMed: 18489733]
- Cosseau, CI; Azzi, A., et al. Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: Critical experimental parameters. *Mol Biochem Parasitol.* 2009; 166(1):70–76. [PubMed: 19428675]
- Curtale G, Citarella F. Dynamic nature of noncoding RNA regulation of adaptive immune response. *Int J Mol Sci.* 2013; 14(9):17347–17377. [PubMed: 23975170]
- Cuthbertson B, Shepard E, et al. Diversity of the penaeidin antimicrobial peptides in two shrimp species. *Immunogenetics.* 2002; 54(6):442–445. [PubMed: 12242595]
- Da Silva TA, Fontes FL, et al. SNPs in DNA repair genes associated to meningitis and host immune response. *Mut Res: Fundam Mol Mech Mutagen.* 2011; 713(1–2):39–47.
- Davidson CR, Best NM, et al. Toll-like receptor genes (TLRs) from *Capitella capitata* and *Helobdella robusta* (Annelida). *Dev Comp Immunol.* 2008; 32(6):608–612. [PubMed: 18164761]
- De Gregorio E, Spellman PT, et al. Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. *Proc Nat Acad Sci.* 2001; 98(22):12590–12595. [PubMed: 11606746]
- DeJong RJ, Emery AM, et al. The mitochondrial genome of *Biomphalaria glabrata* (Gastropoda: basommatophora), intermediate host of *Schistosoma mansoni*. *J Parasitol.* 2004; 90(5):991–997. [PubMed: 15562597]
- Denoeud F, Henriot S, et al. Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science.* 2010; 330(6009):1381–1385. [PubMed: 21097902]
- Dheilily NM, Nair SV, et al. Highly variable immune-response proteins (185/333) from the sea urchin, *Strongylocentrotus purpuratus*: proteomic analysis identifies diversity within and between individuals. *J Immunol.* 2009; 182(4):2203–2212. [PubMed: 19201874]
- Ding SW, Voinnet O. Antiviral Immunity Directed by Small RNAs. *Cell.* 2007; 130(3):413–426. [PubMed: 17693253]
- Dong Y, Cirimotich CM, et al. *Anopheles* NF- κ B-regulated splicing factors direct pathogen-specific repertoires of the hypervariable pattern recognition receptor AgDscam. *Cell Host Microbe.* 2012; 12(4):521–530. [PubMed: 23084919]
- Dong Y, Taylor HE, et al. AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. *PLoS Biol.* 2006; 4(7):e229. [PubMed: 16774454]
- Down TA, Rakyan VK, et al. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotech.* 2008; 26(7):779–785.
- Eddy SR. Non-coding RNA genes and the modern RNA world. *Nat Rev Genet.* 2001; 2(12):919–929. [PubMed: 11733745]
- Eizaguirre C, Lenz TL, et al. Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. *Nat Commun.* 2012; 3:621. [PubMed: 22233631]
- Elango N, Hunt BG, et al. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee. *Apis mellifera Proc Nat Acad Sci.* 2009; 106(27):11206–11211.
- Elango N, Yi SV. DNA methylation and structural and functional bimodality of vertebrate promoters. *Mol Biol Evol.* 2008; 25(8):1602–1608. [PubMed: 18469331]
- Fneich S, Dheilily NM, et al. 5-methyl-cytosine and 5-hydroxy-methyl-cytosine in the genome of *Biomphalaria glabrata*, a snail intermediate host of *Schistosoma mansoni*. *Parasites and Vectors.* 2013
- Freeman JD, Warren RL, et al. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res.* 2009; 19(10):1817–1824. [PubMed: 19541912]

- Galinier R, Portela J, et al. Biomphalysin, a new β Pore-forming toxin involved in *Biomphalaria glabrata* immune defense against *Schistosoma mansoni* PLoS Pathog. 2013; 9(3):e1003216.
- Gavery M, Roberts S. DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*). BMC Genomics. 2010; 11(1):483. [PubMed: 20799955]
- Ghosh J, Lun CM, et al. Invertebrate immune diversity. Dev Comp Immunol. 2011; 35(9):959–974. [PubMed: 21182860]
- Glazov EA, Cottee PA, et al. A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. Genome Res. 2008; 18(6):957–964. [PubMed: 18469162]
- Gowher H, Leismann O, et al. DNA of *Drosophila melanogaster* contains 5-methylcytosine. EMBO J. 2000; 19(24):6918–6923. [PubMed: 11118227]
- Hamon MA, Batsché E, et al. Histone modifications induced by a family of bacterial toxins. Proc Nat Acad Sci. 2007; 104(33):13467–13472. [PubMed: 17675409]
- Harvell CD, Mitchell CE, et al. Climate warming and disease risks for terrestrial and marine biota. Science. 2002; 296(5576):2158–2162. [PubMed: 12077394]
- Hemmrich G, Miller DJ, et al. The evolution of immunity: a low-life perspective. Trends Immunol. 2007; 28(10):449–454. [PubMed: 17855167]
- Hibino T, Loza-Coll M, et al. The immune gene repertoire encoded in the purple sea urchin genome. Dev Biol. 2006; 300(1):349–365. [PubMed: 17027739]
- Ho-Pun-Cheung A, Assenat E, et al. A large-scale candidate gene approach identifies SNPs in SOD2 and IL13 as predictive markers of response to preoperative chemoradiation in rectal cancer. Pharmacogenomics J. 2010
- Hraber PT, Weller JW. On the species of origin: diagnosing the source of symbiotic transcripts. Genome Biol. 2001; 2(9):0037.
- Huang S, Yuan S, et al. Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. Genome Res. 2008; 18(7):1112–1126. [PubMed: 18562681]
- Hudson ME. Sequencing breakthroughs for genomic ecology and evolutionary biology. Mol Ecol Res. 2008; 8(1):3–17.
- Hughes GM, Gang L, et al. Using Illumina next generation sequencing technologies to sequence multigene families in de novo species. Mol Ecol Resour. 2013; 13(3):510–521. [PubMed: 23480365]
- Hutton J, Jegga A, et al. Microarray and comparative genomics-based identification of genes and gene regulatory regions of the mouse immune system. BMC Genomics. 2004; 5(1):82. [PubMed: 15504237]
- Imler JL, Ferrandon D. Le printemps de l'immunité innée couronné à Stockholm Prix Nobel de Médecine 2011 : Bruce A. Beutler, Jules A. Hoffmann et Ralph M. Steinman. Med Sci. 2011; 27:1019–1024.
- Ip JY, Tong A, et al. Global analysis of alternative splicing during T-cell activation. RNA. 2007; 13(4):563–572. [PubMed: 17307815]
- Juranic Lisnic V, Babic Cac M, et al. Dual analysis of the murine cytomegalovirus and host cell transcriptomes reveal new aspects of the virus-host cell interface. PLoS Pathog. 2013; 9(9):e1003611. [PubMed: 24086132]
- Koboldt DC, Chen K, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. 2009; 25(17):2283–2285. [PubMed: 19542151]
- Kondilis-Mangum HD, Wade PA. Epigenetics and the adaptive immune response. Mol Asp Med. 2013 (0).
- Kongchum P, Palti Y, et al. SNP discovery and development of genetic markers for mapping innate immune response genes in common carp (*Cyprinus carpio*). Fish Shellfish Immunol. 2010; 29(2): 356–361. [PubMed: 20420915]
- Kortschak RD, Samuel G, et al. EST Analysis of the Cnidarian *Acropora millepora* Reveals Extensive Gene Loss and Rapid Sequence Divergence in the Model Invertebrates. Current Biology. 2003; 13(24):2190–2195. [PubMed: 14680636]

- Kubinak JL, Ruff JS, et al. Experimental viral evolution to specific host MHC genotypes reveals fitness and virulence trade-offs in alternative MHC types. *Proc Nat Acad Sci.* 2012; 109(9):3422–3427. [PubMed: 22323587]
- Kvennefors ECE, Leggat W, et al. An ancient and variable mannose-binding lectin from the coral *Acropora millepora* binds both pathogens and symbionts. *Dev Comp Immunol.* 2008; 32(12): 1582–1592. [PubMed: 18599120]
- Lafferty KD. The ecology of climate change and infectious diseases. *Ecology.* 2009; 90(4):888–900. [PubMed: 19449681]
- Laurent L, Wong E, et al. Dynamic changes in the human methylome during differentiation. *Genome Res.* 2010
- Lee JH, Ang JK, et al. Analysis and design of RNA sequencing experiments for identifying RNA editing and other single-nucleotide variants. *RNA.* 2013; 19(6):725–732. [PubMed: 23598527]
- Li H, Ruan J, et al. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 2008; 18(11):1851–1858. [PubMed: 18714091]
- Li R, Fan W, et al. The sequence and de novo assembly of the giant panda genome. *Nature.* 2010; 463(7279):311–317. [PubMed: 20010809]
- Li R, Li Y, et al. SNP detection for massively parallel whole-genome resequencing. *Genome Res.* 2009; 19(6):1124–1132. [PubMed: 19420381]
- Login, FdrH; Balmann, Sv, et al. Antimicrobial peptides keep insect endosymbionts under control. *Science.* 2011; 334(6054):362–365. [PubMed: 22021855]
- Loker ES, Adema CM, et al. Invertebrate immune systems - not homogeneous, not simple, not well understood. *Immunological Rev.* 2004; 198(1):10–24.
- Lovoll M, Wiik-Nielsen J, et al. A novel totivirus and piscine reovirus (PRV) in Atlantic salmon (*Salmo salar*) with cardiomyopathy syndrome (CMS). *Virol J.* 2010; 7(1):309. [PubMed: 21067578]
- Lyko F, Foret S, et al. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol.* 2010; 8(11):e1000506. [PubMed: 21072239]
- Lynch KW. Consequences of regulated pre-mRNA splicing in the immune system. *Nat Rev Immunol.* 2004; 4(12):931–940. [PubMed: 15573128]
- Ma F, Xu S, et al. The microRNA miR-29 controls innate and adaptive immune responses to intracellular bacterial infection by targeting interferon- γ . *Nat Immunol.* 2011; 12(9):861–869. [PubMed: 21785411]
- Marth GT, Korf I, et al. A general approach to single-nucleotide polymorphism discovery. *Nat Genet.* 1999; 23(4):452–456. [PubMed: 10581034]
- Martone R, Euskirchen G, et al. Distribution of NF- κ B-binding sites across human chromosome 22. *Proc Nat Acad Sci.* 2003; 100(21):12247–12252. [PubMed: 14527995]
- Mastrangelo AM, Marone D, et al. Alternative splicing: Enhancing ability to cope with stress via transcriptome plasticity. *Plant Sci.* 2012; 185(0):40–49. [PubMed: 22325865]
- McIlroy G, Foldi I, et al. Toll-6 and Toll-7 function as neurotrophin receptors in the *Drosophila melanogaster* CNS. *Nat Neurosci.* 2013; 16(9):1248–1256. [PubMed: 23892553]
- Medzhitov R, Schneider DS, et al. Disease tolerance as a defense strategy. *Science.* 2012; 335(6071): 936–941. [PubMed: 22363001]
- Messier-Solek C, Buckley KM, et al. Highly diversified innate receptor systems and new forms of animal immunity. *Sem Immunol.* 2010; 22(1):39–47.
- Metchnikoff, E. *Lectures on the Comparative Pathology of Inflammation.* N. Y.: 1893. (Reprinted by Dover, 1968), (Translated by F. A. Starling and E. H. Starling)
- Meunier J, Lemoine Fdr, et al. Birth and expression evolution of mammalian microRNA genes. *Genome Res.* 2013; 23(1):34–45. [PubMed: 23034410]
- Miller D, Hemmrich G, et al. The innate immune repertoire in Cnidaria -ancestral complexity and stochastic gene loss. *Genome Biol.* 2007; 8(4):R59. [PubMed: 17437634]
- Minton K. Immune regulation: Long non-coding RNAs in the immune system. *Nat Rev Immunol.* 2013; 13(9):617–617.

- Mitta G, Adema CM, et al. Compatibility polymorphism in snail/schistosome interactions: From field to theory to molecular mechanisms. *Dev Comp Immunol.* 2012; 37(1):1–8. [PubMed: 21945832]
- Modrek B, Resch A, et al. Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nuc Acids Res.* 2001; 29(13):2850–2859.
- Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. *Genomics.* 2008; 92(5):255–264. [PubMed: 18703132]
- Moxon S, Jing R, et al. Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Res.* 2008; 18(10):1602–1609. [PubMed: 18653800]
- Nadal-Ginard B, Smith C, et al. Alternative splicing is an efficient mechanism for the generation of protein diversity: contractile protein genes as a model system. *Adv Enzyme Regul.* 1991; 31:261–286. [PubMed: 1877390]
- Nair SV, Del Valle H, et al. Macroarray analysis of coelomocyte gene expression in response to LPS in the sea urchin. Identification of unexpected immune diversity in an invertebrate. *Physiol Genomics.* 2005; 22(1):33–47. [PubMed: 15827237]
- NCBI. 3556 Eukaryotic genome sequencing projects. 2013. from <http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi>
- Nunez-Acuna G, Gallardo-Escarate C. Identification of immune-related SNPs in the transcriptome of *Mytilus chilensis* through high-throughput sequencing. *Fish Shellfish Immunol.* 2013 (0).
- O'Neill LP, Turner BM. Immunoprecipitation of native chromatin: NChIP. *Methods.* 2003; 31(1):76–82. [PubMed: 12893176]
- Pan Q, Shai O, et al. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet.* 2008; 40(12):1413–1415. [PubMed: 18978789]
- Pankratz VS, Vierkant R, et al. Associations between SNPs in candidate immune-relevant genes and rubella antibody levels: a multigenic assessment. *BMC Immunology.* 2010; 11(1):48. [PubMed: 20923569]
- Pati A, Heath LS, et al. ClaMS: A Classifier for Metagenomic Sequences. *Stand Genomic Sci.* 2011; 5(2):248–253. [PubMed: 22180827]
- Patz JA, Graczyk TK, et al. Effects of environmental change on emerging parasitic diseases. *Int J Parasitol.* 2000; 30(12–13):1395–1405. [PubMed: 11113264]
- Peng Z, Cheng Y, et al. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat Biotech.* 2012; 30(3):253–260.
- Perrin C, Lepesant MJ, et al. *Schistosoma mansoni* Mucin Gene (*SmPoMuc*) Expression: Epigenetic Control to Shape Adaptation to a New Host. *PLoS Pathog.* 2013; 9(8):e1003571. [PubMed: 24009504]
- Putnam NH, Srivastava M, et al. Sea Anemone Genome Reveals Ancestral Eumetazoan Gene Repertoire and Genomic Organization. *Science.* 2007; 317(5834):86–94. [PubMed: 17615350]
- Quinlan AR, Stewart DA, et al. Pyrobayes: an improved base caller for SNP discovery in pyrosequences. *Nat Meth.* 2008; 5(2):179–181.
- Ramani AK, Calarco JA, et al. Genome-wide analysis of alternative splicing in *Caenorhabditis elegans*. *Genome Res.* 2010; 21(2):342–348. [PubMed: 21177968]
- Rast JP, Messier-Solek C. Marine invertebrate genome sequences and our evolving understanding of animal immunity. *Biol Bull.* 2008; 214(3):274–283. [PubMed: 18574104]
- Rast JP, Smith LC, et al. Genomic Insights into the Immune System of the Sea Urchin. *Science.* 2006; 314(5801):952–956. [PubMed: 17095692]
- Regev A, Lamb MJ, et al. The role of DNA methylation in invertebrates: developmental regulation or genome defense? *Mol Biol Evol.* 1998; 15(7):880.
- Reynolds S, Rolff J. Immune function keeps endosymbionts under control. *J Biol.* 2008; 7(8):28. [PubMed: 18947377]
- Roberts SB, Gavary MR. Does DNA methylation facilitate phenotypic plasticity in invertebrates? *Front Physiol.* 2012; 2
- Sabourault C, Ganot P, et al. Comprehensive EST analysis of the symbiotic sea anemone, *Anemonia viridis*. *BMC Genomics.* 2009; 10(1):333. [PubMed: 19627569]

- Sanford JR, Wang X, et al. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* 2009; 19(3):381–394. [PubMed: 19116412]
- Schmucker D, Chen B. Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes. *Genes Dev.* 2009; 23(2):147–156. [PubMed: 19171779]
- Schmucker D, Clemens JC, et al. Drosophila Dscam Is an Axon Guidance Receptor Exhibiting Extraordinary Molecular Diversity. *Cell.* 2000; 101(6):671–684. [PubMed: 10892653]
- Schulenburg H, Kurtz J, et al. Introduction. Ecological immunology. *Phil Trans Roy Soc B Biol Sci.* 2009; 364(1513):3–14.
- Schulte LN, Eulalio A, et al. Analysis of the host microRNA response to Salmonella uncovers the control of major cytokines by the let-7 family. *EMBO J.* 2011; 30(10):1977–1989. [PubMed: 21468030]
- Schweitzer B, Meng L, et al. Immune response biomarker profiling application on ProtoArray® protein microarrays. *The Urinary Proteome.* 2010; 641:243–252.
- Serre D, Lee BH, et al. MBD-isolated genome sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nuc Acids Res.* 2010; 38(2):391–399.
- Sheldon BC, Verhulst S. Ecological immunology: costly parasite defences and trade-offs in evolutionary ecology. *Trends Ecol Evol.* 1996; 11(8):317–321. [PubMed: 21237861]
- Shen Y, Wan Z, et al. A SNP discovery method to assess variant allele probability from next-generation resequencing data. *Genome Res.* 2010; 20(2):273–280. [PubMed: 20019143]
- Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotech.* 2008; 26(10):1135–1145.
- Shimizu TS, Takahashi K, et al. CpG distribution patterns in methylated and non-methylated species. *Gene.* 1997; 205(1–2):103–107. [PubMed: 9461383]
- Shinzato C, Shoguchi E, et al. Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature.* 2011; 476(7360):320–323. [PubMed: 21785439]
- Simpson VJ, Johnson TE, et al. *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. *Nuc Acids Res.* 1986; 14(16):6711–6719.
- Sorber K, Dimon MT, et al. RNA-Seq analysis of splicing in *Plasmodium falciparum* uncovers new splice junctions, alternative splicing and splicing of antisense transcripts. *Nuc Acids Res.* 2011; 39(9):3820–3835.
- Sorek R, Kunin V, et al. CRISPR – a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Micro.* 2008; 6(3):181–186.
- Srivastava M, Simakov O, et al. The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature.* 2010; 466(7307):720–726. [PubMed: 20686567]
- Stahlschmidt ZR, Rollinson N, et al. Are all eggs created equal? Food availability and the fitness trade-off between reproduction and immunity. *Func Ecol.* 2013 n/a-n/a.
- Sullivan CS. New roles for large and small viral RNAs in evading host defences. *Nat Rev Genet.* 2008; 9(7):503–507. [PubMed: 18490927]
- Suzuki MM, Kerr ARW, et al. CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res.* 2007; 17(5):625–631. [PubMed: 17420183]
- Taganov KD, Boldin MP, et al. NF- κ B-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. *Proc Nat Acad Sci.* 2006; 103(33):12481–12486. [PubMed: 16885212]
- Terwilliger D, Buckley K, et al. Distinctive expression patterns of 185/333 genes in the purple sea urchin, *Strongylocentrotus purpuratus*: an unexpectedly diverse family of transcripts in response to LPS, β -1,3-glucan, and dsRNA. *BMC Mol Biol.* 2007; 8(1):16. [PubMed: 17331248]
- Toth AL, Varala K, et al. Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. *Science.* 2007; 318(5849):441–444. [PubMed: 17901299]
- Trapnell C, Pachter L, et al. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009; 25(9):1105–1111. [PubMed: 19289445]
- Vaz C, Ahmad H, et al. Analysis of microRNA transcriptome by deep sequencing of small RNA libraries of peripheral blood. *BMC Genomics.* 2010; 11(1):288. [PubMed: 20459673]

- Vera JC, Wheat CW, et al. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol Ecol*. 2008; 17(7):1636–1647. [PubMed: 18266620]
- Vidal-Dupiol J, Zoccola D, et al. Genes Related to Ion-Transport and Energy Production Are Upregulated in Response to CO₂-Driven pH Decrease in Corals: New Insights from Transcriptome Analysis. *PLoS ONE*. 2013; 8(3):e58652. [PubMed: 23544045]
- Vilcinskas A, Mukherjee K, et al. Expansion of the antimicrobial peptide repertoire in the invasive ladybird *Harmonia axyridis*. *Proc Roy Soc B Biol Sci*. 2013; 280(1750)
- Wang ET, Sandberg R, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008; 456(7221):470–476. [PubMed: 18978772]
- Watson FL, Puttmann-Holgado R, et al. Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science*. 2005; 309(5742):1874–1878. [PubMed: 16109846]
- Wei L, Vahedi G, et al. Discrete Roles of STAT4 and STAT6 Transcription Factors in Tuning Epigenetic Modifications and Transcription during T Helper Cell Differentiation. *Immunity*. 2010; 32(6):840–851. [PubMed: 20620946]
- Weng, N-p; Araki, Y., et al. The molecular basis of the memory T cell response: differential gene expression and its epigenetic regulation. *Nat Rev Immunol*. 2012; 12(4):306–315. [PubMed: 22421787]
- Westermann AJ, Gorski SA, et al. Dual RNA-seq of pathogen and host. *Nat Rev Micro*. 2012; 10(9): 618–630.
- Yamamoto Y, Verma UN, et al. Histone H3 phosphorylation by IKK- α is critical for cytokine-induced gene expression. *Nature*. 2003; 423(6940):655–659. [PubMed: 12789342]
- Yeo GW, Coufal NG, et al. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol*. 2009; 16(2):130–137. [PubMed: 19136955]
- Zhang P, Zhao M, et al. Whole-genome DNA methylation in skin lesions from patients with. *Psoriasis vulgaris J Autoimmunity*. 2013; 41(0):17–24.
- Zhang SM, Loker ES. The FREP gene family in the snail *Biomphalaria glabrata*: additional members, and evidence consistent with alternative splicing and FREP retrosequences. *Dev Comp Immunol*. 2003; 27(3):175–187. [PubMed: 12590969]
- Zhang SM, Loker ES. Representation of an immune responsive gene family encoding fibrinogen-related proteins in the freshwater mollusc *Biomphalaria glabrata*, an intermediate host for *Schistosoma mansoni*. *Gene*. 2004; 341(0):255–266. [PubMed: 15474308]
- Zhang SM, Yong Z, et al. Expression profiling and binding properties of fibrinogen-related proteins (FREPs), plasma proteins from the schistosome snail host *Biomphalaria glabrata*. *Innate Immunity*. 2008; 14(3):175–189. [PubMed: 18562576]
- Zhang Y, Lv J, et al. HHMD: the human histone modification database. *Nuc Acids Res*. 2010; 38(suppl 1):D149–D154.
- Zhuang X, McPhee K, et al. Rapid transcriptome characterization and parsing of sequences in a non-model host-pathogen interaction; pea-Sclerotinia sclerotiorum. *BMC Genomics*. 2012; C7 - 668 13(1):1–19.
- Ziauddin J, Schneider DS. Where does innate immunity stop and adaptive immunity begin? *Cell Host Microbe*. 2012; 12(4):394–395. [PubMed: 23084909]
- Zisoulis DG, Lovci MT, et al. Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. *Nat Struct Mol Biol*. 2010; 17(2):173–179. [PubMed: 20062054]

Figure 1.

Total number of sequenced genome on the 9th of October 2013 as referenced in (NCBI 2013)

Figure 2.

Model of host-symbiont co-evolution to compare the strength (thickness of arrows) of negative (red) and positive (green) components of interactions between symbiont and host that result in specialisation and diversification of immune responses mechanisms in mutualistic and parasitic interactions, respectively.

Figure 3. Scheme showing the workflow from primary sequencing of DNA or RNA to application of NGS in comparative immunology.

Table 1

Comparison of characteristics of Next Generation Sequencing versus traditional Sanger sequencing, considering same budget/effort.

	NGS	Sanger
Cost per millions base	0.07 – 10 \$ <<< 2400 \$	
Read length	35 – 900 bp < = 400–900 bp	
Accuracy	98 – 99.9% < 99.999%	
Amount of template needed / sequence	NGS < < Sanger	
Multiplexing of samples (individuals/treatment) (or simultaneous analysis of multiple samples)	NGS >>> Sanger	
Reconstruction of full length contig	NGS < = Sanger*	
Genome / Transcriptome assembly	NGS >>> Sanger	
Recovery of rare sequences	NGS >> Sanger	
Sampling of unknown sequences	NGS > Sanger	
Representation of members of genefamilies	NGS >> Sanger	
Sampling of variant sequences	“complete”** versus randomly selective	
Recovery of symbiont/pathogen sequences	NGS > Sanger	
Recovery of methylated sequences **	NGS > Sanger	
Protein- nucleic acid interactions **	NGS > Sanger	
Information on expression level	NGS > Sanger	
miRNA profiling	NGS >>> Sanger	
Comprehensive sequence comparison among species	NGS >> Sanger	

* read-lengths for NGS approaching high throughput Sanger sequencing [BASED ON (EST average= ~400bp Sanger, close to 454)]

** “complete” within confines of sequencing bias by different techniques.

*** using appropriately prepared template