2010

The prospects for sequencing the western corn rootworm genome

Nicholas Miller
University of Nebraska-Lincoln, nmiller4@unl.edu

S. Richards
Human Genome Sequencing Center, Baylor College of Medicine

T. W. Sappington
USDA-ARS, CICGRU, Genetics Laboratory, Iowa State University

JOURNAL OF APPLIED ENTOMOLOGY
J. Appl. Entomol.

The prospects for sequencing the western corn rootworm genome

N. J. Miller[1], S. Richards[2] & T. W. Sappington[1]

1 USDA-ARS, CICGRU, Genetics Laboratory, Iowa State University, Ames, 50011 IA, USA
2 Human Genome Sequencing Center, Baylor College of Medicine, Houston, 77030 TX, USA

Keywords
Diabrotica virgifera virgifera, genomics, next-generation DNA sequencing

Correspondence
Nicholas J. Miller (corresponding author), USDA-ARS, CICGRU, Genetics Laboratory, Iowa State University, Ames, 50011 IA, USA.
E-mail: nicholas.miller@ars.usda.gov

Received: June 1, 2009; accepted: August 19, 2009.

doi: 10.1111/j.1439-0418.2009.01448.x

Abstract

Historically, obtaining the complete sequence of eukaryotic genomes has been an expensive and complex task. For this reason, efforts to sequence insect genomes have largely been confined to model organisms, species that are important to human health and representative species from a few insect orders. This situation is set to change as a number of 'next generation' sequencing technologies are making large-scale DNA sequencing both affordable and accessible. Sequencing the genome of the western corn rootworm, *Diabrotica virgifera virgifera*, is likely to become a realistic proposition within the next 2 years. In the meantime, there is an active community of *Diabrotica* geneticists and biologists who are working to assemble the resources that will be needed for a genome sequencing project. A western corn rootworm genome sequence will be an invaluable resource that will facilitate research into the genetics, evolution and ecology of a major pest of maize agriculture in North America and Europe.

Introduction

The first complete sequence of an insect genome was reported less than a decade ago (Adams et al. 2000). Fittingly, this was the genome of the fruit fly *Drosophila melanogaster*, a premier model organism for genetical and molecular studies of biological processes. Since then, the number of insect species for which genome sequencing efforts have been completed or are underway has grown substantially. As well as additional *Drosophila* species, there has been an emphasis on sequencing the genomes of insects that impact human health as vectors of disease. These include several mosquito species, tsetse fly, sand fly and *Rhodnius prolixus*, the vector of Chagas' disease (Robinson et al. 2006). Economically important domesticated insects have not been ignored as the genomes of the silkworm moth, *Bombyx mori* (Mita et al. 2004; Xia et al. 2004; The International Silkworm Genome Consortium 2008) and the honey bee (Honeybee Genome Sequencing Consortium 2006) both have been sequenced.

To date, only one coleopteran genome has been sequenced, that of the red flour beetle, *Tribolium castaneum* (*Tribolium* Genome Sequencing Consortium 2008). The availability of complete genome sequences from four different orders is proving to be an invaluable resource for studying the evolution and fundamental biology of insects. For example, comparative analyses have revealed many genes that were present in the common ancestor of insects and vertebrates but have been lost in the lineage leading to Diptera such as *Drosophila* and mosquitoes (Honeybee Genome Sequencing Consortium 2006, *Tribolium* Genome Sequencing Consortium 2008).

Tribolium was a natural first choice for a beetle genome to sequence. The species is an important pest of stored grain products. It is also an excellent model species, being easy to maintain in culture and amenable to genetic manipulation, and has a

relatively small genome at around 204 Mbp (Brown et al. 1990). These characteristics are typical of the organisms that have been selected for genome sequencing, particularly small genome size. Of 18 arthropod genome projects listed by Robinson et al. (2006), most involve organisms with genomes less than 500 Mbp and only two concern genomes greater than 1 Gbp. This is to be expected. Historically, genome sequencing efforts have been large projects costing millions of US dollars and requiring the involvement of many institutions. Naturally, there has been a tendency to select insect species that, in addition to their interesting biological properties, happen to have small genomes that are less expensive to sequence.

In recent years, a number of technological developments have caused the cost of DNA sequencing to plummet. Consequently, sequencing comparatively large genomes from non-model organisms is becoming a realistic proposition. In this paper, we briefly review current 'next-generation' sequencing methods and improvements to these methods that may be expected in the foreseeable future. We also discuss how these new technologies may enable the genome of the western corn rootworm to be sequenced, a project that although desirable has hitherto been prohibitively expensive.

Why sequence the western corn rootworm genome?

The western corn rootworm, *Diabrotica virgifera virgifera* (Coleoptera: Chrysomelidae) is arguably the most destructive pest of maize in the USA where the annual costs of control and lost yield exceed $1 billion (Sappington et al. 2006). Furthermore, the species is invasive in Europe and has become widespread through a combination of intra-continental movements and recurrent transatlantic introductions (Miller et al. 2005; Ciosi et al. 2008). Because of its pest status, the western corn rootworm is often exposed to strong selection pressures in the form of pest management technologies and methods. Unfortunately, it has demonstrated a remarkable ability to adapt to these selection pressures. Populations in Nebraska evolved resistance to cyclodiene insecticides, used as soil treatments, within a decade (Ball and Weekman 1962). Cyclodienes were succeeded by organophosphate and carbamate insecticides, which also were widely used in Nebraska for both adult and larval control. By the early 1990s, adult resistance to these insecticide classes had evolved (Meinke et al. 1998), largely due to elevated esterase activity (Miota et al. 1998; Scharf et al. 1999; Zhou et al. 2002). As well as evolving resistance to synthetic insecticides, western corn rootworm has adapted to crop rotation as a cultural method of control. Crop rotation exploits the females' strong preference for maize fields as oviposition sites and the univoltine life cycle in which eggs pass through the winter in diapause. Typically in the USA, maize is rotated with soybeans in a 2-year cycle. Because females do not lay eggs in soybean fields, when the fields are rotated to maize the following year, they are free of eggs and, crucially, larvae that are responsible for causing damage. Adaptation to crop rotation was first seen in east-central Illinois (Levine and Oloumi-Sadeghi 1996), where females had lost their ovipositional fidelity to maize fields (Levine et al. 2002; Mabry and Spencer 2003; Pierce and Gray 2006), and has since spread throughout most of Illinois and into several neighbouring states and provinces (Gray et al. 2009). Adaptation to transgenic maize that expresses insecticidal proteins from *Bacillus thuringiensis* has not been observed in the field but selected for in the laboratory (Lefko et al. 2008; Meihls et al. 2008).

There is an active community of researchers focused on the genetics and genomics of western corn rootworm, particularly in the context of its adaptability and invasiveness. The collaborative nature of this community is evidenced by the *Diabrotica* genetics consortium, which comprises 36 research groups in 23 institutions in North America, Europe and Australia (Sappington et al. 2006). There is an ongoing effort to use genetic variation to understand the process by which western corn rootworm is invading Europe (Miller et al. 2005; Ciosi et al. 2008). A variety of approaches have been used to investigate the adaptation to crop rotation including genome scanning (Miller et al. 2007), gene expression studies (Garabagi et al. 2008) and the development of an expressed sequence tag (EST) library (H. M. Robertson, S. T. Ratcliffe, J. Thimmapuram, L. Lin & G. Gong: GenBank accessions EW761110–EW777362). An EST library also has been developed from the larval midgut, primarily as a tool to identify potential targets for insecticides (Siegfried et al. 2005). Efforts are underway to map quantitative trait loci (QTL) associated with resistance to Bt-toxins (K.J. Oswald and M.J. Bagley, pers. comm.) and organophosphate insecticides (B.D. Siegfried, L.J. Meinke, N.J. Miller and T.W. Sappington, unpubl. data). All of these research programs would benefit considerably from a western corn rootworm genome sequence. The most immediate justification for

pursuing a genome sequence is simply economic. It will be cheaper in the long run to sequence the western corn rootworm genome than for individual laboratories to continue developing genomic tools and working on problems piecemeal.

A genome sequence, in combination with a good linkage map would greatly facilitate identifying genes involved in mapped traits. Once a trait has been mapped to a particular genomic region, the genes in that region potentially involved in the trait can be looked up from the genome sequence (Hunt et al. 2007; Ammons and Hunt 2008). A likely early application would be candidate genes at QTL for resistance to insecticides, especially Bt-toxins. A genome sequence would also aid the study of genes that are not easy to identify by other means. Genes that are only expressed transiently, at very low levels and/or in only a few cells are unlikely to be found in EST libraries. For example, genome sequences have been essential to the identification of insect olfactory and gustatory receptor genes (e.g. Clyne et al. 1999; Robertson and Wanner 2006; Engsontia et al. 2008). Equally, genes that are evolving rapidly can be hard to isolate via homology with known genes in other species but still can be identified in genome sequences (e.g. Hill et al. 2002).

DNA sequencing technologies

Regardless of the DNA sequencing technology used, the size of a genome is a key determinant in the cost of sequencing it. In this respect, the western corn rootworm presents a significant challenge because its haploid genome is estimated to be around 2.5×10^9 bp (Sappington et al. 2006). This is comparable to the genomes of mammals such as human (2.9×10^9 bp) or mouse (2.5×10^9 bp). Although the estimate of the western corn rootworm genome size is preliminary (Sappington et al. 2006), it is the only one currently available and therefore provides the best guide to the likely costs of a genome sequencing effort.

All of the complete eukaryotic genomes that have been published to date have been based on Sanger dideoxy termination chemistry and capillary electrophoresis. Since its invention (Sanger et al. 1977), the Sanger method of DNA sequencing has undergone substantial refinement and the probability of errors in the modern version is well understood (Ewing and Green 1998). Consequently, Sanger sequencing is the standard against which other sequencing technologies are compared. Typically, a single Sanger sequencing run, or 'read' will provide up to 800 bases of reliable data.

Eukaryotic genomes have been sequenced following one of two strategies. The more traditional approach is to generate libraries of clones of genomic DNA that are assembled into a physical map of the genome. The individual clones are then sequenced and, because their orientation and relative order is known, the entire genome sequence can be reconstructed (e.g. The *Arabidopsis* Genome Initiative 2000). This strategy has largely been supplanted by 'whole genome shotgun' sequencing, which relies primarily on computer algorithms to assemble unordered sequence reads into large contiguous sequences (e.g. Adams et al. 2000). Typically at least sixfold coverage (i.e. on average, each genomic location is sequenced six times) is needed to obtain a 'draft' genome sequence.

Both the traditional and whole genome shotgun approaches require huge numbers of individual sequencing reactions to be performed and their products analysed by electrophoresis. Consequently, genome sequencing has historically been carried out at dedicated genome sequencing centres that are adequately equipped to cope with the task and are highly automated to maximize throughput. Genome centres also benefit from economies of scale that substantially reduce the unit cost of sequencing reactions. Nevertheless, we estimate that to sequence the western corn rootworm genome to sixfold coverage using Sanger sequencing would cost between 10 and 15 million US dollars. Although there is a strong argument for sequencing the western corn rootworm genome, there are equally strong arguments to support sequencing a multitude of smaller, more tractable genomes from other organisms. It therefore seems improbable that financial support could be obtained to sequence the western corn rootworm genome using standard Sanger methods.

In the past 5 years, several new DNA sequencing methods, often collectively termed 'next-generation' sequencing, have become available. These methods differ substantially in their underlying chemistry and the nature of the data they produce but all can be characterized as generating many short (compared to Sanger chemistry) sequence reads in a massively parallel manner, thus greatly reducing the cost per sequenced nucleotide. The specialized chemistry and engineering required to perform this feat relies on sophisticated instruments, specific to each method and supplied by a single vendor.

The first commercially available next-generation technology was '454' (pronounced 'four-five-four')

pyrosequencing (Margulies et al. 2005), developed by 454 Life Sciences (Branford, CT, USA), now a division of Roche (Basel, Switzerland). Samples are prepared for sequencing by obtaining short DNA fragments, normally by mechanical shearing, to which synthetic adaptors are ligated. Single-stranded template molecules are then captured onto microscopic beads via a bound oligonucleotide that complements the adaptors using a template concentration low enough that at most one molecule is expected to attach to a bead. The single template molecules are then amplified by 'emulsion-PCR' or 'em-PCR'. The em-PCR takes place in an emulsion of oil and PCR reaction mixture such that single beads are held in a droplet of reaction mixture, where the single template on each bead is clonally amplified. After em-PCR, the beads with amplified template are combined with a sequencing primer (complementary to the adaptor sequence) and DNA polymerase and deposited into the wells of a specialized 'Pico Titer Plate'. The size of the wells ensures only one bead can enter a well. Additional, smaller beads carrying ATP sulphurylase and luciferase are then added to the wells. Pyrosequencing by synthesis is performed using a specialized instrument, the 'Genome Sequencer FLX' (Roche). The four nucleotides are sequentially and repeatedly flowed over the wells of the plate along with additional reagents for pyrosequencing. At each cycle, if the nucleotide is incorporated into the growing DNA strand by DNA polymerase, pyrophosphate is released. The bead-bound ATP sulphurylase catalyses the reaction of the pyrophosphate with adenosine phosphosulphate to produce ATP which is then used by luciferase to convert luciferin to oxyluciferin. This last reaction produces light which is recorded by the instrument's optics system. An image of the entire plate is taken at each flow cycle. Thus, many template molecules are sequenced in parallel, each appearing as a 'spot' on the image. The sequence of each template is determined by the nucleotide washes at which light is emitted.

Since its introduction, 454 sequencing technology has been refined, yielding significant improvements in performance. At present, a single sequencing run on the Genome Sequencer FLX instrument generates over 10^6 individual reads at around 400 bases each or up to 6×10^8 bases in total. A further recent improvement is the development of sample preparation methods to enable 'paired-end' reads. These consist of a pair of reads in opposite orientation (i.e. towards each other) separated by a known distance. The availability of such data greatly aids the process of assembling sequence reads into a complete genome and may largely, if not entirely, eliminate the need for supplementary Sanger sequencing to complete *de novo* genome sequencing.

Six instrument runs would be needed for onefold coverage of the western corn rootworm genome. Because individual reads are shorter than for Sanger sequencing, 15-fold coverage would be required to assemble a draft genome sequence. At present day prices, this would cost around $700 000. Given some additional costs associated with obtaining and preparing samples and the bioinformatics needed to assemble the genome, a 454-based effort to sequence the western corn rootworm would cost approximately $1 million. Although this is by no means a trivial sum, it is not much greater than the levels of funding that could be obtained through a single competitive research grant, albeit a generous one. For comparison, the US department of Agriculture Cooperative State Research, Education, and Extension Service (USDA-CSREES) Agriculture and Food Research Initiative Competitive Grants Program's maximum budget for proposals in the Arthropod and Nematode Biology and Management: Tools, Resources and Genomics program is $750 000. It is also encouraging to note that the per-base cost of 454 sequencing declined by about fourfold during 2008. This was largely due to increases in read length and the number of reads per instrument run. Roche has already announced efforts to increase read lengths up to 1000 bases (see press releases available at http://www.454.com). It therefore seems reasonable to suppose that further significant reductions in the costs of 454 sequencing will occur in the near future.

A second next-generation sequencing technology is based on the use of 'reversible terminator chemistry' (Bentley et al. 2008). This approach was commercialized by Solexa Inc., subsequently acquired by Illumina Inc. (San Diego, CA, USA). As with 454 sequencing, the initial stage is to obtain a sample of short DNA fragments, normally by mechanical shearing and to ligate synthetic adaptors to them. DNA molecules are then immobilized by hybridization to oligonucleotides, complementary to the adaptor sequence and bound to the surface of a 'flow cell'. The lawn of oligonucleotides bonded to the flow cell not only acts to capture the DNA templates but also provides the primers for 'bridge amplification' in which the original template molecules are amplified to generate many additional copies, clustered at the site of the original molecule. Thus, the surface of the flow cell is populated with millions of

discrete clusters, each of which generates a sequence read.

Sequencing by the reversible terminator method is performed using an automated instrument, the 'Illumina Genome Analyzer' (Illumina). Inside the instrument, modified nucleotides that carry different fluorophores (corresponding to the four different bases) and a proprietary reversible terminator that blocks further DNA strand extension are added to the flow cell. A DNA polymerase adds the complementary nucleotides to newly synthesized strands in each cluster on the flow cell and the terminator ensures only a single nucleotide is added to each strand. At this point, the fluorophores are excited and an image of the flow cell is taken. The identity of the nucleotide incorporated into each cluster is determined by the wavelength of light emitted by the fluorophore. Next, a chemical treatment is applied that cleaves off the fluorophore and also removes the terminator. The entire process is repeated to determine the next nucleotide in the sequence of each cluster.

The key feature of Illumina's sequencing technology is that it generates enormous numbers of very short reads. Ongoing refinements to the technology have more than doubled the read lengths that can be obtained since it was introduced but the maximum read length is currently only 75 bases. This is offset by the sheer number of clusters that can be sequenced in parallel – up to 1×10^8, yielding as much as 7×10^9 bases of sequence per instrument run. The instrument manufacturer has already announced efforts to increase both read length and density of clusters on the flow cell. Furthermore, methods have been developed to carry out paired-end sequencing which not only eases assembly of the reads but effectively doubles the amount of sequence generated per instrument run although the cost and time taken increase proportionally. Because the individual reads are so short, a much greater depth of coverage, as much as 50-fold is needed for a genome sequence. Even so, the technology is quite affordable. Sequencing a 2.5 Gb genome to 50-fold coverage with Illumina's technology could be done for around $200 000 at current prices.

Sequencing by Oligonucleotide Ligation and Detection (SOLiD) is yet another next-generation sequencing platform, sold by Applied Biosystems (Foster City, CA, USA). SOLiD differs from most other DNA sequencing technologies in that it is based on ligation reactions rather than DNA synthesis. The SOLiD system generates up to 20×10^9 bases of sequence of data per run but individual reads are extremely short, up to 50 bases. Because of these short read lengths, the SOLiD system is not being used for *de novo* sequencing of eukaryotic genomes, as far as we are aware. The most recent next-generation sequencing technology to become available is based on sequencing-by-synthesis and is unusual because the sequencing template does not need to be amplified. Rather, each sequencing read is obtained directly from a single DNA molecule (Harris et al. 2008). This single-molecule sequencing platform is supplied by Helicos Biosciences (Cambridge, MA, USA) and presently offers comparable read lengths and total sequence output to the SOLiD system. Other approaches to next-generation sequencing are being pursued but are presently at the proof-of-concept stage. A particularly interesting example is a method that involves the direct observation of individual DNA polymerase molecules as they add labelled nucleotides to a new DNA strand and which has the potential to generate sequencing reads thousands of bases in length (Eid et al. 2009).

Given the remarkable cost reductions that next-generation sequencing methods offer, it is reasonable to ask why many new genomes have not been sequenced using these technologies. In part, the answer is that the technology has only been available and reasonably mature for a short period of time. Thus, projects to sequence eukaryotic genomes largely or entirely by next-generation methods are not yet complete. Ongoing insect pest projects include *Helicoverpa amigera*, the Hessian fly (*Mayetiola destructor*) and the medfly (*Ceratitis capitata*). A more fundamental issue is that of assembling short read data from next-generation sequencing methods into a complete or even draft genome sequence. The application of next-generation sequencing to large eukaryotic genomes has so far been driven by interest in 're-sequencing' of organisms for which a genome sequence has already been determined by Sanger sequencing, especially human genomes (e.g. Wheeler et al. 2008; Chiang et al. 2009). Under these circumstances, the next-generation reads can be mapped to the existing reference sequence so the issue of assembly does not arise. In contrast, the all against all alignment approach that is the basis of algorithms for whole genome shotgun sequence assembly of Sanger sequence data does not perform well with the shorter read lengths typical of next-generation methods (Quinn et al. 2008). In particular, the read lengths of some platforms are near or less than the minimum alignment length required for positive alignment of two sequence reads, and the sheer numbers of the required short reads

greatly increase the computational effort required. This problem is being mitigated as read lengths increase, particularly for 454 pyrosequencing (Rounsley et al. 2009). A 'hybrid' strategy in which the genome is cloned into a large-insert library that is then assembled into a physical map followed by next-generation sequencing of pools of contiguous clones may also be effective (Rounsley et al. 2009).

There is also strong interest within the bioinformatics community to develop new methods that cope better with *de novo* assembly of next-generation sequencing data and significant advances are being made (Butler et al. 2008; Chaisson and Pevzner 2008; Miller et al. 2008; Zerbino and Birney 2008). Early assembly algorithms focused on very short reads (30–35 bp) and small, relatively repeat free prokaryotic genomes. As read lengths have increased, focus has shifted to larger and more complex eukaryotes. Recent announcements have declared successful *de novo* genome assemblies using only data from next-generation sequencing, for the giant panda (see http://www.illumina.com for press releases) and the oil palm (see http://www.454.com for press releases). It is important to stress that neither of these genome assemblies have yet been published in the peer-reviewed literature. Consequently, it is not known how good the assemblies are or how they were achieved. Nevertheless, these announcements are encouraging and indicate that a purely next-generation approach to genome sequencing is feasible. Overall, it appears that next-generation technologies and their attendant bioinformatic tools are not quite mature enough to be relied upon for a successful attempt to sequence the western corn rootworm genome. However, it seems highly probable that this situation will change within 1–2 years, especially as incrementally increasing read lengths reduce the complexity of the required assemblies.

Prerequisites for a genome sequencing effort

Before any genome sequencing effort can be launched, a number of resources need to be available. In the case of the western corn rootworm, many of these are already available or are being actively developed by members of the *Diabrotica* research community. The first and probably most pressing need is for a better estimate of the size of the genome. The amount of sequencing that must be done is highly dependent on the genome's size. The preliminary estimate of the western corn rootworm's genome size, 2.5×10^9 bp (Sappington et al. 2006), is not reliable enough to accurately estimate the total cost of sequencing. Fortunately, work has already begun to obtain an accurate size estimate by means of flow cytometry (B.S. Coates, pers. comm.).

In addition to an estimate of the size of the genome, a preliminary evaluation of its sequence composition and organization will be important. Given its large size, we may be certain that a considerable proportion of the western corn rootworm genome is made up of repetitive elements, which will pose problems during genome assembly. An assessment of the size and nature of the repetitive elements that are common in the western corn rootworm genome will give an indication of how severe the assembly problems might be. Ideally, an awareness of the difficulties likely to be encountered due to repetitive elements will lead to strategies that help to mitigate them. Furthermore, the relationship between coding genes and repetitive elements is important. If western corn rootworm genes tend to have large introns that contain repetitive elements with high sequence identity (as seen in the maize genome) then genome assembly will be especially challenging. If on the other hand, repetitive elements and coding genes are partitioned into different genome regions it may be relatively easy to assemble at least the more interesting gene-rich regions. Work is already underway that will provide the needed preliminary genome sequence data. As part of a larger project to provide genomic resources for western corn rootworm, we (N.J. Miller and T.W. Sappington with B.D. Siegfried and H.M. Robertson) are in the process of sequencing up to 100 Bacterial Artificial Chromosome clones each of which contains a large section (average of 105 kb) of the western corn rootworm genome. The sequences of these BACs will be deposited in a web-accessible database that is being developed as part of the same project. A further asset being developed as part of this project is a high-density linkage map of the western corn rootworm genome that will aid in the high-level assembly of a genome sequence.

Unfortunately, the *Tribolium* genome sequence will probably not be of great assistance for assembling a western corn rootworm genome. Because the two species are in different superfamilies (Tenebrionoidea and Chrysomeloidea), they are sufficiently unrelated that synteny (evolutionary conservation of gene-order along chromosomes) is unlikely. This suspicion is supported by the more than 10-fold difference in genome size between the two species. However, the availability of the *Tribolium* annotated gene set will greatly enhance the automated annotation of the western corn rootworm genome. This is because protein sequences (and their underlying

genes) tend to be more evolutionarily conserved than large-scale genome organization. Consequently, *Tribolium* can provide a set of relatively closely related protein sequences that can be used to 'train' automated gene prediction software. This will be especially important for detecting genes that are unique to the Coleoptera. The known exon order of *Tribolium* genes could also be used to correct gene-scale assembly errors in the western corn rootworm genome sequence that may occur due to the presence of very long introns.

Another important resource will be the development of an inbred western corn rootworm strain that can be used as a source of DNA for sequencing. Ideally, the DNA sample used in a genome sequencing effort would contain no allelic variation. Such variation complicates the assembly process as the assembly algorithm may be unable to distinguish between reads that are similar but non-identical due to allelic variation, sequencing errors and paralogues. The latter are similar sequences from different genomic locations that have arisen due to gene duplications in the organism's evolutionary past. Realistically, allelic variation cannot be eliminated from a western corn rootworm strain but it can be minimized by successive generations of inbreeding. Recently, USDA scientists working at Brookings, South Dakota have begun work to develop an inbred strain (B.W. French, pers. comm.). The strain is being derived from a laboratory colony that was subjected to artificial selection to remove the obligatory diapause that the eggs undergo (Branson 1976). Consequently, it is possible to rear four, rather than one, generations per year. This strain has already lost roughly a quarter of its natural variation compared to wild populations (Kim et al. 2007). Multiple inbred lines are being developed, each maintained through a cross between a single pair of siblings at each generation. The redundancy of developing multiple lines is important because it is likely that many will go extinct as a result of inbreeding depression. It is anticipated that the final inbred strain will pass through up to eight generations of sib mating before being maintained as a normal laboratory colony that can be made available to the western corn rootworm research community.

In this paper, we have concentrated on the prospects for sequencing the western corn rootworm genome. Once a genome sequence is obtained, an arguably larger challenge presents itself, that of genome annotation. Annotation is the process of identifying and describing features such as coding genes, regulatory elements and mobile elements within the genome sequence. Much of this can be done computationally but 'manual' annotation by scientists is also important. The National Center for Biotechnology Information (NCBI) offers to work in partnership with research communities to annotate genome sequences. An important resource for annotation is ESTs, sequences that are derived from messenger RNA as these can be used to identify coding genes within the genomic sequence. An appreciable number of ESTs have already been obtained for the western corn rootworm (see above). Nevertheless, an expanded EST set covering additional tissues, life stages and biological states will be an important asset. Next-generation sequencing is likely to make the acquisition of new ESTs both inexpensive and rapid.

Conclusion

Although a project to sequence the western corn rootworm genome is not yet quite feasible, the rapid pace of developments in sequencing technology mean that it almost certainly will be within 2 years. In the meantime, the resources needed for a successful genome sequencing effort can be, and are being developed, in particular:

1 An inbred strain.
2 A good estimate of the size of the western corn rootworm genome.
3 Preliminary data on the sequence organization of the western corn rootworm genome from bacterial artificial chromosome clones.
4 A high-density linkage map for western corn rootworm.
5 Expanded coverage of EST data for western corn rootworm.

Acknowledgements

This study was developed from discussions held during the Western Corn Rootworm Genome Sequencing Workshop at the Entomological Society of America's Annual Meeting in Reno, Nevada in November 2008. We thank all of the participants in the workshop who generously shared their knowledge and insight. Two anonymous reviewers provided constructive criticism of an earlier draft of the manuscript.

References

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA,

Galle RF et al., 2000. The genome sequence of *Drosophila melanogaster*. Science 287, 2185–2195.

Ammons AD, Hunt GJ, 2008. Identification of quantitative trait loci and candidate genes influencing ethanol sensitivity in honey bees. Behav. Genet. 38, 531–553.

Ball HJ, Weekman GT, 1962. Insecticide resistance in the adult western corn rootworm in Nebraska. J. Econ. Entomol. 55, 439–441.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR et al., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456, 53–59.

Branson TF, 1976. The selection of a non-diapause strain of *Diabrotica virgifera* (Coleoptera: Chrysomelidae). Entomol. Exp. Appl. 19, 148–154.

Brown SJ, Henry JK, Black WC IV, Denell RE, 1990. Molecular genetic manipulation of the red flour beetle: genome organization and cloning of a ribosomal protein gene. Insect Biochem. 20, 185–193.

Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB, 2008. ALLPATHS: de novo assembly of whole-genome shotgun microreads. Genome Res. 18, 810–820.

Chaisson MJ, Pevzner PA, 2008. Short read fragment assembly of bacterial genomes. Genome Res. 18, 324–330.

Chiang DY, Getz G, Jaffe DB, O'Kelly MJT, Zhao X, Carter SL, Russ C, Nusbaum C, Meyerson M, Lander ES, 2009. High-resolution mapping of copy-number alterations with massively parallel sequencing. Nat. Meth. 6, 99–103.

Ciosi M, Miller NJ, Kim KS, Giordano R, Estoup A, Guillemaud T, 2008. Invasion of Europe by the western corn rootworm, *Diabrotica virgifera virgifera*: multiple transatlantic introductions with various reductions of genetic diversity. Mol. Ecol. 17, 3614–3627.

Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR, 1999. A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. Neuron 22, 327–338.

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B et al., 2009. Real-time DNA sequencing from single polymerase molecules. Science 323, 133–138.

Engsontia P, Sanderson AP, Cobb M, Walden KK, Robertson HM, Brown S, 2008. The red flour beetle's large nose: an expanded odorant receptor gene family in *Tribolium castaneum*. Insect Biochem. Mol. Biol. 38, 387–397.

Ewing B, Green P, 1998. Base-calling of automated sequencer traces using phred II. Error probabilities. Genome Res. 8, 186–194.

Garabagi F, French BW, Schaafsma AW, Pauls KP, 2008. Increased expression of a cGMP-dependent protein kinase in rotation-adapted western corn rootworm (*Diabrotica virgifera virgifera* L.). Insect Biochem. Mol. Biol. 38, 697–704.

Gray ME, Sappington TW, Miller NJ, Moeser J, Bohn MO, 2009. Adaptation and invasiveness of western corn rootworm: intensifying research on an a worsening pest. Annu. Rev. Entomol. 54, 303–321.

Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J, DiMeo J, Efcavitch JW et al., 2008. Single-molecule DNA sequencing of a viral genome. Science 320, 106–109.

Hill CA, Fox AN, Pitts RJ, Kent LB, Tan PL, Chrystal MA, Cravchik A, Collins FH, Robertson HM, Zwiebel LJ, 2002. G protein-coupled receptors in *Anopheles gambiae*. Science 298, 176–178.

Honeybee Genome Sequencing Consortium, 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 443, 931–949.

Hunt G, Amdam G, Schlipalius D, Emore C, Sardesai N, Williams C, Rueppell O, Guzmán-Novoa E, Arechavaleta-Velasco M, Chandra S et al., 2007. Behavioral genomics of honeybee foraging and nest defense. Naturwissenschaften 94, 247–267.

Kim KS, French BW, Sumerford DV, Sappington TW, 2007. Genetic diversity in laboratory colonies of western corn rootworm (Coleoptera : Chrysomelidae), including a nondiapause colony. Environ. Entomol. 36, 637–645.

Lefko SA, Nowatzki TM, Thompson SD, Binning RR, Pascual MA, Peters ML, Simbro EJ, Stanley BH, 2008. Characterizing laboratory colonies of western corn rootworm (Coleoptera: Chrysomelidae) selected for survival on maize containing event DAS-59122-7. J. Appl. Entomol. 132, 189–204.

Levine E, Oloumi-Sadeghi H, 1996. Western corn rootworm (Coleoptera: Chrysomelidae) larval injury to corn grown for seed production following soybeans grown for seed production. J. Econ. Entomol. 89, 1010–1016.

Levine E, Spencer JL, Isard S, Onstad DW, Gray ME, 2002. Adaptation of the western corn rootworm to crop rotation: evolution of a new strain in response to a management practice. American Entomologist 48, 94–107.

Mabry TR, Spencer JL, 2003. Survival and oviposition of a western corn rootworm variant feeding on soybean. Entomol. Exp. Appl. 109, 113–121.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y, Chen Z et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.

Meihls LN, Higdon ML, Siegfried BD, Miller NJ, Sappington TW, Ellersieck MR, Spencer TL, Hibbard BE, 2008. Increased survival of western corn rootworm on transgenic corn within three generations of

on-plant greenhouse selection. Proc. Natl. Acad. Sci. U S A 105, 19177–19182.
Meinke LJ, Siegfried BD, Wright RJ, Chandler LD, 1998. Adult susceptibility of Nebraska western corn rootworm (Coleoptera : Chrysomelidae) populations to selected insecticides. J. Econ. Entomol. 91, 594–600.
Miller N, Estoup A, Toepfer S, Bourguet D, Lapchin L, Derridj S, Kim KS, Reynaud P, Furlan L, Guillemaud T, 2005. Multiple transatlantic introductions of the western corn rootworm. Science 310, 992.
Miller NJ, Ciosi M, Sappington TW, Ratcliffe ST, Spencer JL, Guillemaud T, 2007. Genome scan of *Diabrotica virgifera virgifera* for genetic variation associated with crop rotation tolerance. J. Appl. Entomol. 131, 378–385.
Miller JR, Delcher AL, Koren S, Venter E, Walenz BP, Brownley A, Johnson J, Li K, Mobarry C, Sutton G, 2008. Aggressive assembly of pyrosequencing reads with mates. Bioinformatics 24, 2818–2824.
Miota F, Scharf ME, Ono M, Marcon P, Meinke LJ, Wright RJ, Chandler LD, Siegfried BD, 1998. Mechanisms of methyl and ethyl parathion resistance in the western corn rootworm (Coleoptera: Chrysomelidae). Pestic. Biochem. Physiol. 61, 39–52.
Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanamori H, Namiki N, Kitagawa M, Yamashita H, Yasukochi Y et al., 2004. The genome sequence of silkworm, *Bombyx mori*. DNA Res. 11, 27–35.
Pierce CMF, Gray ME, 2006. Western corn rootworm, *Diabrotica virgifera virgifera* LeConte (Coleoptera : Chrysomelidae), oviposition: a variant's response to maize phenology. Environ. Entomol. 35, 423–434.
Quinn N, Levenkova N, Chow W, Bouffard P, Boroevich K, Knight J, Jarvie T, Lubieniecki K, Desany B, Koop B et al., 2008. Assessing the feasibility of GS FLX pyrosequencing for sequencing the Atlantic salmon genome. BMC Genomics 9, 404.
Robertson HM, Wanner KW, 2006. The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. Genome Res. 16, 1395–1403.
Robinson GE, Evans JD, Maleszka R, Robertson HM, Weaver DB, Worley K, Gibbs RA, Weinstock GM, 2006. Sweetness and light: illuminating the honey bee genome. Insect Mol. Biol. 15, 535–539.
Rounsley S, Marri P, Yu Y, He R, Sisneros N, Goicoechea J, Lee S, Angelova A, Kudrna D, Luo M et al., 2009. De novo next generation sequencing of plant genomes. Rice 2, 35–43.
Sanger F, Nicklen S, Coulson AR, 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U S A 74, 5463–5467.
Sappington TW, Siegfried BD, Guillemaud T, 2006. Coordinated *Diabrotica* genetics research: accelerating progress on an urgent insect pest problem. American Entomologist 52, 90–97.
Scharf ME, Meinke LJ, Siegfried BD, Wright RJ, Chandler LD, 1999. Carbaryl susceptibility, diagnostic concentration determination, and synergism for U.S populations of western corn rootworm (Coleoptera: Chrysomelidae). J. Econ. Entomol. 92, 33–39.
Siegfried BD, Waterfield N, ffrench-Constant RH, 2005. Expressed sequence tags from *Diabrotica virgifera virgifera* midgut identify a coleopteran cadherin and a diversity of cathepsins. Insect Mol. Biol. 14, 137–143.
The Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796–815.
The International Silkworm Genome Consortium, 2008. The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. Insect Biochem. Mol. Biol. 38, 1036–1045.
Tribolium Genome Sequencing Consortium, 2008. The genome of the model beetle and pest *Tribolium castaneum*. Nature 452, 949–955.
Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen Y, Makhijani V, Roth GT et al., 2008. The complete genome of an individual by massively parallel DNA sequencing. Nature 452, 872–876.
Xia Q, Zhou Z, Lu C, Cheng D, Dai F, Li B, Zhao P, Zha X, Cheng T, Chai C et al., 2004. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). Science 306, 1937–1940.
Zerbino DR, Birney E, 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.
Zhou XG, Scharf ME, Parimi S, Meinke LJ, Wright RJ, Chandler LD, Siegfried BD, 2002. Diagnostic assays based on esterase-mediated resistance mechanisms in western corn rootworms (Coleoptera : Chrysomelidae). J. Econ. Entomol. 95, 1261–1266.