Proposal to Sequence the *Nasonia* Genome


Contact Person: J.H. Werren, Phone: 585-275-2394, Fax: 585-275-2070, email: werr@mail.rochester.edu

Overview:

The Parasitic Hymenoptera is a diverse and extremely important insect group. Known also as parasitoid wasps, these insects are natural enemies of a broad range of arthropods of medical, veterinary and agricultural significance. Adult parasitoids are free-living insects that lay their eggs in or on various life stages of other arthropods (e.g. insects, ticks, mites), subsequently killing the host. Most arthropods are attacked by multiple parasitoid species that specialize in different host life stages (e.g. egg, larva, pupa). As a result, parasitoids are major regulators of arthropod populations in nature.

More insects beneficial to humans occur in the Parasitic Hymenoptera than in any other insect group. Parasitoids attack significant vectors of human disease, such as house flies, roaches and ticks (Quicke 1997). They are also extremely important regulators of agricultural pests, and as such have a major impact on human nutrition. The realized and potential impact of these insects on human health, by reducing destruction of food, is truly staggering. For example, in the US alone biological control programs using parasitoid wasps save approximately $20 billion dollars annually in crop loss to newly invasive species (see Letter, K. Hackett); and this does not include the vast savings from biological control of native pests. The knowledge gained from full genome sequencing of a parasitoid will likely lead to methods for further enhancing their utility for control of disease vectors and agricultural pests.

Parasitoids have incredibly diverse life histories and modes of development (Godfray 1994, Quicke 1997). For example, early development ranges from a syncytial blastoderm to a holoblastic cleavage more similar to what is found in mammals (Grbic et al 1998). Parasitoids have evolved a variety of mechanisms for manipulating host immunity, physiology and behavior through their venoms (Quicke 1997, Rivers et al 1999). Among their effects, venoms can induce temporary paralysis, selective apoptosis, alterations in lipid uptake, and host immune suppression. Given the incredible diversity of parasitoids (over 170,000 species) and host associations, this represents an impressive untapped pharmacopoeia, access to which will be improved by full sequencing of a parasitoid genome.

Parasitic Hymenoptera have a form of sex determination, called haplodiploidy, which makes them particularly suited for genetic studies. Females are diploid and develop from fertilized eggs, whereas males are haploid and develop parthenogenetically from unfertilized eggs. This form of sex determination allows geneticists to exploit many of the advantages of haploid genetics in an otherwise complex eukaryotic organism. As a result, parasitoids and in particular those in the genus *Nasonia*, are emerging as models for studies of complex genetic traits, because such traits are readily dissected genetically by taking advantage of male haploidy.

For these and other reasons, obtaining a full genome sequence for a parasitoid is likely to yield many benefits for human health and our understanding of important biological processes. The parasitoid wasp *Nasonia vitripennis* is the logical choice for this effort. *Nasonia* has been the subject of genetic research for over 50 years and is the model for parasitoid genetics (Whiting 1967). It is an extremely tractable laboratory organism with an active research community. Therefore, information from the genome sequence will immediately be utilized in experimental study. The biology of *Nasonia* is better understood than any other parasitic hymenopteran. As a result, there is broad consensus within the parasitoid and biological control communities for selection of *Nasonia* as the parasitoid of choice for full genome sequencing (see Letters).
A. Specific Biological Rationales For The Nasonia Sequence Data

The genus Nasonia is a complex of three closely related species of parasitic insects. The biological rationales for sequencing of the Nasonia genome fall into four basic categories: (a) Nasonia and close relatives are parasitoids of houseflies and other filth flies, which are major vectors of human and veterinary disease, (b) Nasonia is the genetic model for all parasitic wasps – as such a full genome sequence will provide an impetus for enhancing the use of these important insects for control of harmful arthropods of medical and nutritional significance, (c) Nasonia is emerging as a model for complex genetic traits – the inherent genetic advantages of this system could provide a new and economically efficient tool for gene discovery of complex traits affecting human health and development, and (d) Nasonia is well positioned for comparative genomic studies that will enhance the utility of it and other more basal insects for informing the human genome, relative to the more derived genomes of drosophilid flies. More detail on these and other points is provided below.

1. Suitability of Nasonia as an experimental system.


1.1 Basic biology and advantages of Nasonia as an experimental system: Nasonia consist of three closely related parasitoid species, N. vitripennis (Nv), N. giraulti (Ng), and N. longicornis (Nl). These small insects, about _-_- the size of D. melanogaster, sting and lay their eggs upon the pupae of various flies, notably blow flies, flesh flies and house flies. Nasonia are incredibly easy to work with in the laboratory. They have a short generation time (2 weeks at 25 C) and large family sizes (over 500 offspring per female, typically 20-50 per blow fly host). Individuals are easily sexed in the immobile pupal stage, and adults do not readily fly allowing handling without anesthetization. Pupae and adults can also be maintained in the refrigerator for 1-2 months. Strains of Nasonia can be induced to enter a larval diapause, which permits storage under refrigeration for up to 1.5 years without maintenance. This allows the storage of large numbers of strains. Visible and molecular markers are available in Nasonia, as well as additional genetic resources (see below). Hosts are available commercially and easy to rear.

Normally, the three Nasonia species are genetically isolated from each other due to Wolbachia, an intracellular bacterium that causes sperm-egg incompatibilities (Werren 1997). However, these bacteria are easily eliminated by antibiotics – the species are then interfertile and genes and genetic regions can readily be moved between them by backcrossing (e.g. Weston et al 1999). There are 5 chromosomes in Nasonia (with visible markers on each) and a high level of synteny between the species based on recombination maps and preliminary cytology of hybrids.

Nasonia is the “lab rat” of the Hymenoptera – the most tractable and genetically developed organism in this insect order. In fact, it has inherent advantages as a genetic system that rivals even the current model insect systems that have been much more developed. It uniquely combines in a complex eukaryotic system the features of haploid genetics, short generation time (2 weeks), laboratory tractability, and availability of closely related inter-fertile species. Haploid genetics facilitates whole genome mutant screens for any trait of interest. For example, it made relatively easy a large scale genetic screen for mutations affecting early development that has identified 1/3 of genes equivalent to early development genes found in Drosophila (Pultz et al 2000). Haploid genetics also facilitates rapid mapping of genes, and characterization of gene interactions involved in complex traits. The presence of closely related interfertile species provides a wealth of molecular markers for fine-scale mapping and positional cloning of any gene of interest. Due to haplodiploidy, lethal mutants tightly linked to genes of interest anywhere in the genome can be readily generated. These are used to quickly reduce the size of regions that have been introgressed from one species to the other to under 1 centimorgan, which greatly facilitates positional cloning of regions/genes of interest. Unlike the honeybee, Nasonia has a genetic form.
of sex determination that permits inbreeding and it readily inbreeds both in nature and in the lab without the deleterious affects found in many diploid organisms. This permits the generation of very healthy highly inbred strains because harmful recessives are purged in the haploid males.

1.2 Resources and technologies: In addition to the inherent advantages of this system for research, genetic/genomic tools are rapidly advancing in this system. Among the resources currently available in *Nasonia* are a mutant linkage map containing over 20 visible mutants, RAPD, Microsat and AFLP linkage maps (approximately 300 molecular markers with the number rapidly growing), an 11X coverage BAC library (average insert size 111KB), a lambda genomic library, chromosome specific dissection libraries, and two embryonic cDNA libraries. An EST project is underway and it has verified that both cDNA libraries are of good quality in terms of insert size and complexity (D. Leaf). Parental RNAi has recently been successfully used in *Nasonia*, and a transformation system is under development (C. Desplan). TiLLing (Target Induced Local Lesions in Genome) is planned for *Nasonia* (J. Werren, see Comai letter). A battery of hybrid inbred lines are available between Nv/Ng, Nv/Nl and Nl/Ng. These can be exploited for mapping of interspecies differences in behavior, development and physiology.

1.3 Mutation screens: Due to haplodiploidy, the complete genome of *Nasonia* can be efficiently screened for mutations in the F2 generation because recessive mutations anywhere in the genome are immediately apparent (e.g. Pultz and Leaf 2003). Males are mutagenized and crossed to control females. The F1 females are then set as virgins – they produce haploid F2 males that are screened for phenotypes of interest. F1 females can also then be mated to maintain recessive lethal or sterile mutations (Saul et al 1965).

1.4 Positional Cloning: *Nasonia* is an excellent system for positional cloning – particularly for complex genetic traits. By exploiting the closely related inter-fertile species, an abundance of molecular markers are available for fine-scale mapping and positional cloning. Other important features are a short generation time, high recombination rate, and ease of generating linked visible and lethal markers to assist in recombinant walking. Thus, it is well positioned for going from Quantitative Trait Loci (QTL’s) to genes. Such an effort is underway now to clone QTL involved in wing cell size differences between Nv and Ng. Males of two species differ in wing cell size. Genes regulating cell growth and size are of medical significance, since mutations in them are often associated with human cancer (Nicholson and Anderson 2002). Several cell growth regulators (*pten, s6K, InR, Tor, pi3k*) have been mapped in *Nasonia* using interspecies RFLP differences (see figure). Haploid sons from F1 females of an interspecies cross (Nv x Ng) were used to map chromosomal regions affecting wing cell size (wm and set). As seen in the figure, major QTL and non-additive (epistatic) interactions among QTL (indicated by arrows) are readily detected in the haploid males, because they are not obscured by dominance interactions (Gadau et al 2002). Three major QTL from Ng have now been backcrossed into Nv and linked visible and lethal mutations have been used to reduce the size of the introgressed regions to (in one case) less than 0.5 cm (approximately 175 Kb). Such genes are prime candidates for positional cloning. Similar approaches are underway for behavioral genes (Beukeboom letter) and can be applied to a wide range of traits, utilizing the related species, strains or selection lines.
2. Improving human health

2.1 Control of Human Disease Vectors: The importance of *Nasonia* for control of human disease vectors falls into two categories (a) *Nasonia* and its relatives are parasitoids used in biological control of the common house fly, a major vector of human disease, and (b) *Nasonia* is the genetic model for all parasitic Hymenoptera, which as a group are biological control agents for a wide range of invertebrate vectors of human disease. Therefore, both for the specific control of the house fly, and more generally for control of invertebrate vectors of disease, full genome sequencing of *Nasonia* will benefit human health.

The common house fly (*Musca domestica*) and its relatives (commonly referred to as filth flies) are disease vectors of major significance to human health (Greenberg 1971). The house fly is known to harbor over 100 human pathogens, including those responsible for coxsackie disease, conjunctivitis, and many enteric diseases (cholera, salmonellosis, shigellosis, coli dysentery, amoebic dysentery). Evidence indicates that it is a vector for many of these diseases. Houseflies contribute to the spread of diarrhea, which is a major source of morbidity and mortality in developing nations.

*Nasonia* and it close relatives (e.g. *Muscidifurax*, *Spalangia* and *Trichomalopsis*) are biological control agents of *Musca* (Patterson and Rutz 1986, Legner 1995). These parasitoids are sold routinely by commercial insectaries for control of houseflies and other filth flies in dairy farms, feedlots and poultry rearing operations (Axtel and Rutz 1986). A full genome sequence of *Nasonia* will provide data useful in genetic modification of these insects to make them much more effective control agents of these serious pests. A specific example is host preference – a major host preference gene has recently been mapped in *Nasonia* (Werren unpublished). Identification of this gene by positional cloning will be a major step forward in eventual manipulation of these and other parasitoid insects for biological control.

More generally, many important vectors of human disease are attacked by parasitic wasps, which play a role in regulating their numbers. These disease vectors include cockroaches (food borne diseases), tsetse fly (sleeping sickness), sandflies (leishmania), ticks (Lyme disease, rickettsioses), and reduviid bugs (chagas disease) (Quicke 1997). Parasitoids are routinely sold for biological control of cockroaches (e.g. [http://www.rinconvitova.com/cockroach_program.htm](http://www.rinconvitova.com/cockroach_program.htm)), and cockroaches are significant passive vectors of food-borne disease. The genome sequence of *Nasonia* will provide baseline information for improvement of parasitoids for biological control of arthropod vectors of human disease.

2.2 Human nutrition: Human nutrition is impacted negatively by the large losses to crops and domestic animals caused by pest insects. It is estimated that 15 – 20 percent of crops are destroyed by insects before being utilized as food (Pimentel 1997). Furthermore, the nutritional value of plants and animals stressed by insect damage declines even at pest densities well below economic injury levels. Parasitoid wasps are arguable the most important agents in biological control of arthropods of agricultural importance. In the United State alone, invasive pests cost this country over $130 billion annually. To counter these realized costs and anticipated threats, natural biological control by use of indigenous and introduced parasitoid wasps has been a key tool (Pimentel 1997, see Hackett letter). *N. vitripennis* would be the first genome sequence of a parasitoid. A genome sequence would help us to better understand crucial biological features of parasitoids like host finding, sex ratio, host preference and egg laying rates all of which are important to make parasitoid more efficient in biological control. Ultimately it might lead to a method to design parasitoids for specific pests.

There is a large international community of parasitoid wasp researchers working on different species - we estimate over 100 research laboratories in the United States alone. Many of these researchers will be able to exploit the *Nasonia* genome in their studies, in particular to assist in gene discovery and development of genomic resources for other parasitoids used in biological control and basic research.

2.3 Pesticides and Biological Control: The use of biological pest control systems also benefits human health because their application leads to a significant reduction of pesticides. Pesticides exposure, both acute and chronic, is associated with a wide range of negative effects on human health. These include neurological dysfunction and disease, infertility and delayed sexual maturation, reduced birth weight, immunosuppression, and breast cancer (Kamel and Hoppin
The long-term health consequences of these chemicals are now becoming more widely understood (for more examples see NIH online journal Environmental Health Perspectives [http://ehp.niehs.nih.gov/]). The extensive use of parasitoid wasps in biological control programs reduces human exposure to these harmful agents. Although the extent of reduction has not been assessed, it is likely to be large given the major economic gains (which have been assessed) from use of biological control agents (e.g. letters from Bolckmans, Godfray, Hackett, van Lenteren, Strouthamer). Sequencing the *Nasonia* genome will be a major step towards harnessing parasitoids for even more effective biological control of pest insects, for example by manipulating host preference, environmental tolerances, et cetera. Therefore, this information could eventually lead to a reduction in pesticide use, and its negative health consequences.

### 2.4 Complex Genetic Traits & Human Disease

A major argument for sequencing of *Nasonia* is that it is an emerging model for the analysis of complex genetic traits. Many human diseases involve complex interactions among multiple genes and environmental factors. Complex diseases such as **cancer**, **heart disease**, **diabetes**, and **some forms of mental illness** are the most important and widespread causes for death and illness in humans (Altmuller et al. 2001). The underlying genetic architecture of these diseases involves interactions among many genes of varying effect and genetic interactions with environmental factors.

Nearly 100 human disease genes have been identified by positional cloning. However, these successes have been mainly limited to rare disorders and genes of very large effect. Discovery of genes involved in diseases with more complex genetic influences has been much more difficult. Among the reasons are the high costs of large-scale linkage analyses in humans, complex gene-environment interactions, and the difficulty of associating markers with phenotypes in genetically variable populations (Altmuller et al 2001). Therefore, there is a need for tractable model systems for analysis of complex genetic traits – permitting the discovery of candidate genes that then can be focused on in mammalian and human studies. *Nasonia* is an ideal emerging system for this approach, because of its unique features (e.g. male haploidy, ease of handling and rearing, closely related interfertile species) that are particularly suited for complex genetic trait analysis.

The study of complex traits of medical importance in *Nasonia* can help to decipher the genetic architecture of such traits and might ultimately lead to a better diagnosis and treatment of complex diseases. The basic approach in *Nasonia* is as follows: (a) taking advantage of male haploidy, genes affecting complex traits (e.g. alcohol tolerance, metabolic rate, toxin sensitivity) can be rapidly mapped and gene interactions uncovered. This can be accomplished using strains, selection lines or species that differ in the phenotype of interest; (b) individual genes can then be isolated and cloned by positional means using recombination to flanking lethals, or visible and molecular markers. The approach uses our ability to efficiently and inexpensively produce large numbers of recombinants; and finally (c) environment interactions with genes can be readily investigated by taking advantage of the ability to produce highly inbred lines and/or hundreds of genetically identical females in the F3 generation.

Thus, *Nasonia* can be used as an efficient system for gene discovery of traits of medical importance, to then permit more focused (and therefore efficient and cost effective) studies of candidate genes in humans. A full genome sequence of *Nasonia vitripennis* is absolutely essential for the potential of this system for complex trait gene discovery to be realized. It will permit the rapid movement from Quantitative Trait Loci (QTL) mapping to candidate gene identification.

### 2.5 Pharmacology of Parasitoid Venoms

Parasitoids provide an incredibly rich and unutilized pharmacopoeia of venoms that may serve as leads for developing new classes of synthetic chemical insecticides and therapeutic compounds affecting cell physiology. The venoms of parasitic wasps are known to have a wide range of effects on their hosts, including arrestment of development, alteration in growth and physiology, suppression of immune responses, induction of paralysis, oncosis or apoptosis, and alteration of host behavior (Jervis and Copeland, 1996; Quicke, 1997). Parasitoid venoms can elicit these diverse host responses either through paralytic or non-paralytic means. Parasitoid venoms (including those of *Nasonia*) trigger hyperlipemia (Nakamatsu and Tanaka, 2003; Rivers and Denlinger, 1994a,b, 1995), block host molting through venom disruption of hormone receptors (Coudron et al., 1994), or suppress host immune
responses either alone (Rivers et al., 1999, 2002) or synergistically with endosymbiotic viruses (polydnavirus, entomopox virus) (Guzo and Stoltz, 1985; Strand and Noda, 1991). Of particular significance to human health are recent findings that Nasonia venom is highly toxic to several mosquitoes that are vectors of such diseases as malaria, encephalitis, yellow fever, and West Nile (Rivers unpubl results). Such biological and physiological diversity suggests that these venoms contain a wealth of compounds with unique chemistries and modes of action. Undoubtedly, these venoms target tissues and receptors unexploited by insecticides or drugs currently on the market, increasing their potential use as tools for enhancing human health. Parasitic Hymenoptera are among the most speciose of insect groups, with at least 170,000 to 600,000 species (LaSalle and Gauld 1991). Given this incredible diversity, the potential pharmacological resource is immense. The Nasonia genome will provide the basis for studying diversity of venoms in different parasitoid species.

3. Informing Human Biology & Providing Additional Surrogate Systems:

3.1 Comparative Genomics: Nasonia is well positioned phylogenetically to assist in identifying orthologs of important genes in insects and mammals, and combines this with a genetically tractable system for functional analysis. Because Drosophila have a rather derived genome for a number of features, Nasonia may be useful for annotation and functional analysis of vertebrate genes not present or too rapidly evolving in Drosophila. Surprisingly, a significant number of genes in the Honey Bee genome had orthologs with the human genome rather than the Drosophila genome. It is likely that the Nasonia genome will provide additional orthologs to the human genome. Preliminary data from the Nasonia EST project (D. Leaf) is consistent with this view. Excluding highly conserved genes (e.g. histones and ribosomal proteins), ten percent of the Nasonia EST's have significant similarity to human genes but no counterpart in Drosophila. Examples include matches to a pancreatic cDNA, elongation protein zeta (involved in axon guidance), and anaphase promoting complex subunit 13 (involved in cell cycle regulation). An additional 25 percent of Nasonia EST's are more similar to human genes than are their Drosophila counterparts (based on a 10^6 fold improvement in blast score) which reflects the derived nature of many Drosophila genes, whereas 10 percent are significantly less similar to human orthologs than are their Drosophila counterparts. These data, although preliminary, support the view that the Nasonia genome will have human homologs that are not present in Drosophila or that are much more similar than the Drosophila counterpart. Given that Nasonia is a very tractable genetic and experimental system (e.g. efficient full genome mutant screening, analysis of gene interactions), it should offer new avenues for analysis of homologous human genes in an insect system.

On a higher level these genomes will help us to understand features such as regulatory domain evolution, frequency and type of non-coding DNA, and metabolic capabilities. The mechanisms of how enzymes and pathways evolve in central regulatory cascades, and cis-regulatory evolution can be analyzed by a number of bioinformatics tools using the Nasonia genome (Dandekar and Saureborn 2002, Gadau and Dandekar letter, Clark letter). Because we propose partial coverage of the two sibling species Ng and Nl, examples of rapid cis-regulatory evolution are likely to emerge.

3.2 Aging – Nasonia is a good model for aging. The haploid males are short-lived (4-7 days), permitting selection experiments and aging studies in a relatively short time scale. Recently, selection lines have been produced that differ by nearly 25% in average longevity - QTL analyses of these lines (not yet done) will reveal candidate loci for positional cloning, and potential epistatic interactions among loci. Similarly, genes known to affect aging in other organisms (e.g. pten) have been backcrossed between Nasonia species, and results suggest an effect on longevity. Furthermore, interspecies crosses have revealed a strong influence of mitochondrial haplotype on survival in interspecies crosses (Breeuwer and Werren 1995, Velthuis unpublished). The system has potential for augmenting studies of aging in other species, particularly in the ability to quickly uncover gene interactions and gene x mitochondrial interactions. Nasonia also has a larval diapause that results in arrested development, changes in physiology, increased cold tolerance, reduced metabolism, and prolonged lifespan. This has been well studied and the basic cues involved in diapause induction are known (Schneidermann and
Horrwitz 1958, Whiting 1967). Nasonia strains differ in diapause tendency, allowing a genetic dissection of this trait.

3.3 Chromosome abnormalities: Major terminal deletions can result in chromosomes containing only the centromere and some flanking genes. These centric fragments, also referred to as “marker” chromosomes, are found in cancers and implicated in some human developmental abnormalities (Li et al. 2000). Chromosomes with major terminal deletions are easily screened for and maintained in Nasonia. This is accomplished by using linked visible or molecular markers and screening for recovery of the maternally-derived marker in haploid males (e.g. Beukeboom et al 1993, Perrot-Minnot and Werren 2001, Perfectti and Werren 2001). Mitotic and meiotic stability and chromosome “healing”, presumably by telomere acquisition, can be studied in the system in a whole organism context. The supernumerary Paternal Sex Ratio chromosome (Werren, 1991) causes destruction of the other sperm chromosomes (but not itself) in fertilized eggs in Nasonia. A full genome sequence will help in identifying genes involved in this disorder and will be informative about chromosome processing mechanisms and early fertilization processes in general.

4. Expanding our understanding of basic biological processes:

4.1 Developmental Biology: Nasonia is proving an excellent organism for addressing a number of outstanding issues in the cellular processes immediately following fertilization, including pronuclear fusion. This is a fascinating event as the paternal and maternal genomes must get acquainted and share a common nucleus. Little is known about this critical time in development because it is difficult to observe in most genetically tractable organisms. For example, although Drosophila has proven an excellent organism for the cellular studies, because the large size of the egg the initial events following fertilization occur deep within the yolky interior of the embryo making it difficult to image these in real-time. In addition, these events proceed very rapidly following fertilization making it very difficult to obtain and prepare a sample in time to capture the event in real-time. The smaller Nasonia embryos and the slower pace of development have made Nasonia an excellent system for live analysis of these early events. Already, this has been successfully employed to analyze the mechanisms of pronuclear migration and centrosome inheritance in Nasonia (Tram and Sullivan 2000, 2002).

The system is also relevant to comparative studies of embryonic development. Drosophila is a “long germ-band” insect (referring to the fact all segments appear at the same time and occupy most of the egg) while many insects, such as Tribolium, undergo a process of “short germ-band” embryogenesis, in which cells that will give rise to the actual embryo are only a small portion of the egg and are usually localized at the extreme posterior pole (Handel et al. 2002). Tribolium has been successfully developed as a comparative model for insect development. However, Nasonia possesses a number of features that make it a critical additional organism for understanding the evolution of development. Like Drosophila, Nasonia undergoes long germband development, and thus, the basic fate maps of the wasp and fly are very similar, despite their phylogenetic distance. Expression patterns, mutants, and RNAi phenotypes of Nasonia genes are much more directly comparable to those of Drosophila than are Tribolium, which has a very different fate map. The phylogenetic position of Nasonia is also intermediate between Coleoptera and Diptera and will fill a hole between our knowledge of development in these species.

The power of Nasonia genetics has begun to be brought to bear on its early development (Pultz et al 2000, Pultz and Leaf 2003). Taking advantage of its haplo-diploid sex determination, a genetic screen for mutations resulting in embryonic lethality was performed (Pultz et al, 2000). The screen detected genes equivalent to 1/3 of genes identified in Drosophila (e.g. gap, pair rule, segment polarity, homeotic), while others do not fall into any clear class. There appears to be a much increased reliance on zygotic gene expression for anterior-posterior patterning in the wasp in comparison to Drosophila (Pultz, et al., 1999). The stage is now set for a more comprehensive analysis of early development using genetic mutants, RNAi, and cytological approaches.

4.2 Reproduction & Sex Determination: Nasonia is currently the best genetic system for the study of parthenogenetic reproduction – development from unfertilized eggs. Males develop parthenogenetically. Studies in Nasonia have revealed cytological mechanisms by which the absence of centrosomes arriving from the sperm is compensated for (Tram and Sullivan 2000). Understanding the mechanisms of parthenogenesis has immense implications, both for revealing the necessary modifications of cell cycle and development, and for its implications in using
parthenogenesis to enhance economic rearing of biological control agents by producing more of the sex (females) that kill pests. A genome sequence will help in identifying candidate genes and in utility of RNAi and microarrays to study mechanisms of parthenogenesis (see Sullivan letter).

Sex determination in *Nasonia* provides an interesting contrast to those of the honeybee and other insects. Although both the honeybee and *Nasonia* are haplodiploid, in honeybees, ants and many other hymenoptera, sex is determined by heterozygosity at a single locus. Embryos that are hemizygous or homozygous develop into males and heterozygous embryos develop into females. Hence, inbreeding in these species results in production of (homozygous) diploid males that are usually infertile. *Nasonia* has circumvented this problem and inbreeds readily without diploid male production. Mechanisms may involve genomic imprinting (Dobson and Tanoye 1998), and genetic variation can be exploited for study (e.g. Nur et al 1988, Beukeboom unpubl).

### 4.3 Neurobiology & Behavior

*Nasonia* has a long history in behavioral ecology and its behavior was a focus almost from the beginning of its career as a model organism (see Whiting 1967 for early references). It has been the subject of studies of courtship behavior, mate preference, aggression, host finding, host preference, sex ratio adjustment, and learning (e.g. Werren 1980, Olla and King 2000, Beukeboom and Assem 2002, Velthuis et al 2004). Due to the importance of the olfactory system in host finding, *Nasonia* may also prove to be a good model for research on chemoreception and olfaction. Preliminary experiments (J. Gadau) have shown that the brain of *Nasonia* contains large antennal lobes with a relatively large number of well-developed olfactory glomeruli. The use of novel neuroanatomical and neurophysiological techniques like calcium imaging on genetic lines differing in preference, is an extremely promising avenue to study the physiological and genetic architecture of the olfactory system. Various studies have demonstrated remarkable similarities among insect and mammalian olfactory systems, including the one in humans. This is true for odorant receptors in peripheral sensory neurons as well as for the functional and anatomical organization of primary olfactory centers in the brain. Therefore, unraveling odorant receptor genes in relation to functional organization of the central olfactory pathway in species with striking olfactory behaviors, as *Nasonia*, is likely to contribute to our general understanding of functional principles of olfaction.

### 4.4 Bacterial-Insect Interactions

*Nasonia* and *Drosophila* are currently the primary models for studies of insect interactions with the intracellular bacterium *Wolbachia*. *Wolbachia* are an incredibly widespread group of bacteria that infect reproductive tissues and alter cell biology and reproduction in 20 -70 percent of insect species. They are important in terms of the mechanisms of altering host cell cycle and development, evolutionary implications, and in possible biological control applications of these bacteria (Werren 1997). The best molecular cytological studies of *Wolbachia* induced cytoplasmic incompatibility (CI - a sperm-egg incompatibility resulting in improper condensation of paternal chromosomes in the fertilized egg) have been conducted in *Nasonia*. This is due to the clarity of eggs and ease of conducting real-time cytology (Tram and Sullivan 2000, 2002). In addition, *Nasonia* is the prime candidate for mutant screens of host genes interacting with *Wolbachia*, because CI is normally 100% (Breeuwer and Werren 1990), making mutant screens of rescue efficient. In addition, the genetic background of the different *Nasonia* species affects the pattern and level of paternal chromosome destruction due to CI (Bordenstein et al. 2003, Tram et al, in prep.); relevant genes can be backcrossed between the species for analysis. The *Nasonia* genome will permit more efficient positional cloning of the genes involved in these processes.

### 4.5 Evolutionary Genetics

A *Nasonia* genome sequence can help to answer some fundamental questions in evolutionary genetics. For example due to the production of viable and fertile hybrids this species complex can be used to analyze “speciation genes” and the genetic architecture of adaptation (e.g. Weston et al 1999, Gadau et al 1999, 2002). Second, *Nasonia* has been a model for studies of sex ratio evolution for over 25 years (e.g Werren 1980, Nur et al 1988, Shuker and West 2004); uncovering the genetic mechanisms of sex ratio control exercised by these insects will be a major step forward in our understanding of the genetics of adaptation (Shuker and West 2004, and see West letter).

*Nasonia* is approximately 120 MY diverged from Honeybee, in a second major branch of the Hymenoptera (Parasitica). In terms of phylogenetic position, a *Nasonia* genome will be very useful for assigning sequence changes along phylogenetic branches between Honeybee and other
insects. This will be relevant for assigning polarity to changes leading to social insect groups versus parasitoids. Addition of the Nasonia genome will also be informative for assigning specific changes along branches leading to the Hymenoptera, Lepidoptera, Diptera and Coleoptera, thus contributing to our understanding of the evolution of holometabolous insects.

B. Strategic issues in acquiring new sequence data:

1. The demand for the new sequence information: Scientific communities that will immediately utilize information from the Nasonia genome include researchers with interests in the following areas: Nasonia, Parasitoid Biology, Biological Control, Honeybee and Social Insects, Comparative Genomics, and Development (see letters of support). The Nasonia community consists of over 15 active laboratories, and is growing rapidly. Because of recent advances in genetic tools in Nasonia and on-going efforts for positional cloning of genes involved in development, cell growth, and behavior, the genome information will immediately be utilized to assist in positional cloning of biologically important genes, for assisting in RNAi studies, and for comparative genomics of Nasonia relative to other insects and vertebrates.

   The Parasitoid research community is very large, consisting of several hundred laboratories around the world (at least 100 laboratories in the United States alone). There is consensus among leading parasitoid and biological control researchers that Nasonia is the logical first choice for sequencing of a parasitoid genome (e.g. letters Godfray, Hackett, Heraty, Hunter, Quicke, Stouthamer, Strand, van Lenteren). Information from the Nasonia genome will be used to identify important genes in parasitoid biology, as previously described (see support letters). In addition, there is broad interest in utilizing the Nasonia genome to identify and annotate genes involved in important biological processes, from olfaction to behavior to enzymatic pathway structure (e.g letters, Robinson, Evans, Gadau & Dandekar, W. Hunter, Rueppell)

2. The rationale for the complete sequence: Nasonia is emerging as a model for genetics of complex traits, development and evolution. A full genome sequence will allow the system to be exploited as an efficient system for positional cloning of genes affecting diverse phenotypes relevant to human health and basic biology. We propose a moderate 6X coverage of N. vitripennis plus low 1X coverage each for the two sibling species, N. giraulti and N. longicornis. Partial sequencing of the two sibling species will allow us to fully exploit the genetic power of this system for positional cloning, analysis of complex traits, and for comparative genomics. Specifically, it will provide a wealth of sequence markers that can be used for ultra-fine scale mapping and positional cloning, as well a identification of genes and regulatory domains undergoing differential rates of evolution in the three species.

3. Practical issues for sequencing the Nasonia genome:

3.1 Size of the Genome and Readiness for Sequencing: Nv has a modest genome size of 330 – 360 MB. This has been estimated by two independent methods, and (1) Giemsa staining (Rasch et al. 1977), (2) fluorescent flow cytometry (S. Johnston, unpubl). Based on chromosomal studies and linkage maps, the other two species are expected to have very similar genome sizes. Nasonia inbreeds readily without deleterious effects common in outbred diplo-diploid species. Therefore the complexities of sequencing from an outbred strain are avoided. We will sequence the highly inbred AsymCX strain of Nv, which was used for construction of the BAC library. Gram quantities of insects are available for DNA preparation. Although Nasonia is a parasite of species of large flies, very clean material can be obtained using dark pupae, which have extruded gut contents and are effectively bacteria-free. Nasonia pupae are very easy to clean (having a thin smooth pupal skin). High molecular weight DNA can be obtained from this stage, and was used in the BAC library construction. The AsymCX line is cured of the non-obligatory endosymbiont Wolbachia.

3.2 Basic Sequencing Strategies: The Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC) has strong interest in sequencing the Nasonia genome (Gibbs letter and participation of S. Richards in this whitepaper). BCM-HGSC has nearly completed sequencing of the honeybee, and this experience has prepared us for a second hymenopteran. We anticipate no significant problems with the sequencing and assembly of Nasonia. Our approach will be pure whole genome shotgun (WGS) sequencing and assembly from subclone and fosmid libraries.
produced at the BCM-HGSC, which was proven to be effective and to produce random clone coverage of the honeybee genome.

Lessons from the sequencing of the honeybee will be applied to *Nasonia*. The honeybee genome project had problems with the BAC library having incomplete coverage of the genome. Fosmid libraries generated from sheared genomic DNA samples and small insert plasmid libraries were found to give complete genome coverage. Because of these incomplete coverage issues *Nasonia* will be sequenced using a pure whole genome strategy, only using BAC end sequences for scaffolding information necessary to increase scaffold sizes into the Mb range. Nevertheless, preliminary hybridization analysis of the current *Nasonia* EcoR1 BAC library (111Kb ave insert size) indicates that it has good coverage with no major gaps. The BAC end sequencing will be performed in a stepwise manner with an initial sampling of 3,000 BAC clones (1X clone coverage) to further check the random coverage of this library before proceeding to end sequence a total of 30,000 BAC clones to provide a maximum of 10X clone coverage for good scaffolding information over the genome. In addition, 35K clones will be fingerprinted for assembly into contigs. BAC ends of contigs can then be readily mapped onto the 5 *Nasonia* chromosomes using F2 haploid males from Nv x Ng crosses (e.g. Gadau et al 2001, 2002). If the BAC library is found to provide uneven representation of the genome we will end sequence additional fosmid clones to increase scaffolding information sufficient for the genome sequence. As the vast majority of the reads will be short insert pUC18 subclones from sheared genomic DNA, the contig coverage will be fully random as was the case for the honeybee screened sequencing libraries.

Assembly of the honeybee whole genome shotgun sequence has been straightforward using the ATLAS whole genome assembly tools developed at the BCM-HGSC for the assembly of the rat, *D. pseudoobscura*, honey bee, sea urchin and numerous bacteria. Based on our experience we expect no problems assembling a small genome the size of *Nasonia*. The annotation of the honeybee genome sequence was done in collaboration with the EBI using the Ensemble pipeline. This pipeline has now been installed at the BCM-HGSC and tuned for the honey bee; the annotation pipeline will be fine-tuned for the second hymenopteran.

The BCM-HGSC is constantly improving and re-evaluating its sequencing pipeline for the reduction of costs. Additionally we have been evaluating the possibilities of new sequencing technologies as they emerge. One such technology is 454 Inc’s (Branford, CT) massively parallel pyro sequencing. 454 Inc is currently producing a 1X sequence of the *D. pseudoobscura* genome in collaboration with BCM-HGSC to allow evaluation of this new technology, and it’s integration into insect sequencing as an initial step. We will closely follow and evaluate the possibility of using this technology with *Nasonia* if the scientific requirements can be fulfilled.

3.3 Possible Biological Features Affecting Sequencing: *Nasonia* has a relatively simple genome structure, with 5 chromosomes and a relatively high recombination rate (approximately 350KB/cm). The G/C content is 43.1% based on sequencing of 100 BAC ends. Screening the BAC library indicates that it has good coverage. There are no obvious impediments to full genome sequencing of *Nasonia*, and it appears to be highly suitable for this effort. High molecular weight DNA is obtainable from isogenic lines, as evidenced in construction of the BAC library. The Baylor Genome Center will very soon initiate sequencing of 2 BAC clones to provide preliminary assessment of genome complexity. In addition 1000 BAC ends may soon be sequenced to provide baseline information on scaffolds for genome assembly.

3.4 Other possible sources for (partial) funding of the project: If sequencing of the *Nasonia* genome is approved on scientific merit, then Kevin Hackett (Senior National Program Leader for Biological Control and Insect Genomics, ARS, USDA) will request support of the project by supplemental funding from the USDA (see Hackett letter). NSF provided funds for construction of the BAC library, which will be used in genome assembly. Funds to develop EST’s and a mixed species (Nv+Ng) microarray have been requested as part of a larger proposal on insect genomics (PI: J. Romero-Severson). The proposal was just evaluated and received an outstanding rating, and is highly likely for funding. The EST’s will be used to assist assembly of the *Nasonia* genome.
References


