ASCARIS: THE NEGLECTED PARASITE

Edited by

CELIA HOLLAND
Trinity College, University of Dublin, Dublin
Contents

Foreword xiii
Contributors xv
Introduction xvii

I

BIOLOGY OF ASCARIS

1. Immunology of Ascaris and Immunomodulation
PHILIP J. COOPER AND CAMILA A. FIGUIEREDO

Introduction 4
Immunology of Ascariasis 5
Immune Modulation in Animal Models 8
Evidence for Clinically Relevant Immune Modulation by Ascariasis During Natural Infection 9
Conclusion 12
References 13

2. Ascaris and Allergy
LUIS CARABALLO

Introduction 22
Basics Concepts of the Allergic Response: how Similar is it to the Immune Response to Ascaris? 23
Allergic Diseases in Tropical Regions 25
Ascariasis, Allergic Sensitization, and Allergy Symptoms 26
Ascariasis Influences the Prevalence of Asthma 27
Can Helminth Infections Increase the Allergic Response? 28
The Allergens of A. lumbricoides 30
Ascariasis Influencing the Diagnosis of Allergy 33
Ascariasis and Asthma Severity 34
Parasite Infections, Allergy, and the Hygiene Hypothesis 35
Allergy as a Protective Factor for Ascariasis: the other Face of Co-Evolution 36
The Possible Therapeutic Impact of Ascaris Immunomodulatory Products on Allergy 39
Conclusion 40
References 41
3. Ascaris — Antigens, Allergens, Immunogenetics, Protein Structures
MALCOLM W. KENNEDY

Introduction 52
Stage-Specific Surface and Secreted Antigens 52
Allergens 54
Correlation Between IgE Antibody to ABA-1 and Acquired Immunity 55
Immunogenetics and Unpredictable Immune Repertoires 57
Polymorphisms in Ascaris Antigens 60
Polymorphisms in Parasite Antigens and Concomitant Immunity 63
Innate Immunity and Complement 64
The Structure and Function of the Major Allergen of Ascaris — ABA-1 65
An Unusual Lipid Binding Protein in the Perivitelline Fluid of Ascaris Eggs 70
Concluding Remarks 73
Acknowledgments 74
References 74

4. Implications of Ascaris Co-infection
FRANCISCA ABANYIE AND TRACEY J. LAMB

Introduction 82
General Considerations for Co-Infection Interactions in Ascaris lumbricoides-Infected Individuals 82
Immune Responses to Ascaris lumbricoides 85
Effects of Ascaris on Systemic Co-Infections 87
Hypothetical Effects of Localized Co-Infections on Ascaris Co-Infection 92
Concluding Remarks 97
Acknowledgments 98
References 98

II
MODEL SYSTEMS

5. Larval Ascariasis: Impact, Significance, and Model Organisms
CELIA V. HOLLAND, JERZY M. BEHNKE, AND CHRISTINA DOLD

Introduction 108
The Hepato-Tracheal Migration in Ascariasis 108
The Impact of Larval Migration in Natural Hosts 109
Animal Models of Larval Ascariasis 111
Mice as Model Organisms for Early Ascaris Infection 112
The Mouse as a Model of Ascaris Aggregation 114
# III

## EPIDEMIOLOGY OF ASCARIASIS

7. *Ascaris lumbricoides*: New Epidemiological Insights and Mathematical Approaches  
MARTIN WALKER, ANDREW HALL, AND MARÍA-GLORIA BASÁNEZ  

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>156</td>
</tr>
<tr>
<td>Diagnosis of Infection</td>
<td>161</td>
</tr>
<tr>
<td>Transmission in Communities</td>
<td>163</td>
</tr>
<tr>
<td>Mathematical Approaches</td>
<td>170</td>
</tr>
<tr>
<td>Conclusions</td>
<td>187</td>
</tr>
<tr>
<td>References</td>
<td>189</td>
</tr>
</tbody>
</table>

8. Genetic Epidemiology of *Ascaris*: Cross-transmission between Humans and Pigs, Focal Transmission, and Effective Population Size  
CHARLES D. CRISCIONE  

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>203</td>
</tr>
<tr>
<td><em>Ascaris</em> Cross-Transmission Between Humans and Pigs</td>
<td>205</td>
</tr>
<tr>
<td>Landscape Genetics as a Means to Infer <em>Ascaris</em> Transmission</td>
<td>208</td>
</tr>
<tr>
<td>within a Host Population</td>
<td></td>
</tr>
<tr>
<td>Effective Population Size: Epidemiological Utility and Estimation</td>
<td>211</td>
</tr>
<tr>
<td>Concluding Remarks</td>
<td>224</td>
</tr>
<tr>
<td>References</td>
<td>226</td>
</tr>
</tbody>
</table>

9. Transmission Dynamics of *Ascaris lumbricoides* — Theory and Observation  
T. DÉIRDRE HOLLINGSWORTH, JAMES E. TRUSCOTT, AND ROY M. ANDERSON  

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>232</td>
</tr>
<tr>
<td>What are Mathematical Models Used for?</td>
<td>234</td>
</tr>
</tbody>
</table>
IV
HOST AND PARASITE GENETICS

10. From the Twig Tips to the Deeper Branches: New Insights into Evolutionary History and Phylogeography of Ascaris
MARTHA BETSON, PETER NEJSUM, AND J. RUSSELL STOTHARD

Introduction 266
Evolutionary History of Ascaridoidea and Host Potential 266
Species Concepts 269
Application of Biochemical and Molecular Tools 271
Experimental Studies 272
Insights into A. lumbricoides/suum complex 273
Global and Local Genetic Diversity 274
Evolutionary History and Spread of Ascaris 275
Implications for Control: A Focus on Zanzibar 278
Concluding Remarks and Future Studies 279
Acknowledgments 280
References 280

11. Decoding the Ascaris suum Genome using Massively Parallel Sequencing and Advanced Bioinformatic Methods — Unprecedented Prospects for Fundamental and Applied Research
AARON R. JEX, SHIPING LIU, BO LI, NEIL D. YOUNG, ROSS S. HALL, YINGRUI LI,
PETER GELDHOF, PETER NEJSUM, PAUL W. STERNBERG, JUN WANG,
HUANMING YANG, AND ROBIN B. GASSER

Introduction 288
Advanced Methodologies Established to Sequence, Assemble, and Annotate the Genome and Transcriptomes of A. suum 290
Salient Characteristics of the Genome and Transcriptomes of A. suum: Toward Understanding the Molecular Landscape of the Parasite and the Design of New Interventions 297
Conclusions, Implications, and Major Prospects for Fundamental and Applied Research 307
Acknowledgments 310
References 310
12. Genetics of Human Host Susceptibility to Ascariasis
SARAH WILLIAMS-BLANGERO, MONA H. FENSTAD, SATISH KUMAR, AND JOHN BLANGERO

Introduction 316
Epidemiology of Ascariasis: The Role of Nonrandom Clustering 317
Measuring the Extent of Host Genetic Variation in Infection 318
Functional Genetic Variation is the Source of Heritability 319
Assessing Heritability in Human Pedigrees 320
Identifying Specific Genes Responsible for Heritability of Ascaris Burden 321
Quantitative Genetics Studies of Ascariasis 323
Genome Scans for Genes Influencing Ascaris Infection 324
Further Association Mapping of the Chromosome 13 QTL Using
High Density SNPs 326
Sequencing of TNFSF13B as a Gene Underlying the QTL for Ascaris Infection 327
Other Candidate Gene Studies of Ascariasis 329
More Complex Models: Host/Worm Genetic Effects and Spatial
Factors Influencing Host Ascaris Burden 331
Future of Genetic Research on Ascaris Infection: Deep Sequencing
to Identify Rare Functional Variants 333
Acknowledgments 336
References 336

V

CLINICAL ASPECTS AND PUBLIC HEALTH

13. Ascaris lumbricoides and Ascariasis: Estimating Numbers
Infected and Burden of Disease
SIMON J. BROOKER AND RACHEL L. PULLAN

Introduction 343
The Global Limits of Transmission 346
A Shrinking Global Distribution 346
Global Numbers at Risk and Infected 349
Global Disease Burden 352
Burden of Ascariasis in 2010 358
Limitations of the DALY Approach and a Broader
View of Impact 359
Acknowledgments 359
References 360

14. Impact of Ascaris suum in Livestock
STIG MILAN THAMSBORG, PETER NEJSUM, AND HELENA MEJER

Introduction 363
Ascaris suum and Other Helminths in Pig Production Systems 364
Contributors

Francisca Abanyie  Emory University School of Medicine, Emory Children’s Center, Atlanta, GA, USA
Roy M. Anderson  Imperial College London, London, UK
María-Gloria Basáñez  Imperial College London, London, UK
Jerzy M. Behnke  University of Nottingham, Nottingham, UK
Martha Betson  The Royal Veterinary College, London, UK
John Blangero  Texas Biomedical Research Institute, San Antonio, TX, USA
Simon J. Brooker  London School of Hygiene and Tropical Medicine, London, UK
Luis Caraballo  University of Cartagena, Colombia
Philip J. Cooper  Liverpool School of Tropical Medicine and Hygiene, Liverpool, UK and Pontificia Universidad Católica del Ecuador, Quito, Ecuador
Charles D. Criscione  Texas A&M University, College Station, TX, USA
Christina Dold  University of Oxford, Oxford, UK
Mona H. Fenstad  Norwegian University of Science and Technology, Trondheim, Norway
Camila A. Figuieredo  Universidade Federal da Bahia, Salvador, Brazil
Albis Francesco Gabrielli  World Health Organization, Geneva, Switzerland
Robin B. Gasser  The University of Melbourne, Parkville, Victoria, Australia
Peter Geldhof  Ghent University, Belgium
Andrew Hall  University of Westminster, London, UK
Ross S. Hall  The University of Melbourne, Parkville, Victoria, Australia
Celia V. Holland  Trinity College, Dublin, Ireland
T. Deirdre Hollingsworth  University of Warwick, Coventry, UK and Liverpool School of Tropical Medicine, Liverpool, UK
Aaron R. Jex  The University of Melbourne, Parkville, Victoria, Australia
Malcolm W. Kennedy  University of Glasgow, Glasgow, Scotland, UK
Satish Kumar  Texas Biomedical Research Institute, San Antonio, TX, USA
Tracey J. Lamb  Emory University School of Medicine, Emory Children’s Center, Atlanta, GA, USA
Bo Li  BGI-Shenzhen, Shenzhen, PR China
Yingrui Li  BGI-Shenzhen, Shenzhen, PR China
Shiping Liu  BGI-Shenzhen, Shenzhen, PR China
Aaron G. Maule  Queen’s University Belfast, Belfast, UK
Helena Mejer  University of Copenhagen, Denmark
Antonio Montresor  World Health Organization, Geneva, Switzerland
Peter Nejsum  University of Copenhagen, Denmark
Rachel L. Pullan  London School of Hygiene and Tropical Medicine, London, UK
Lorenzo Savioli  World Health Organization, Geneva, Switzerland
Paul W. Sternberg  California Institute of Technology, Pasadena, CA, USA
J. Russell Stothard  Liverpool School of Tropical Medicine, Liverpool, UK
Antony O.W. Stretton  University of Wisconsin-Madison, Madison, WI, USA
Stig Milan Thamsborg  University of Copenhagen, Denmark
James E. Truscott  Imperial College London, London, UK
Johnny Vlaminck  Ghent University, Belgium
Martin Walker  Imperial College London, London, UK
Jun Wang  BGI-Shenzhen, Shenzhen, PR China
Sarah Williams-Blangero, PhD  Texas Biomedical Research Institute, San Antonio, TX, USA
Huanming Yang  BGI-Shenzhen, Shenzhen, PR China
Neil D. Young  The University of Melbourne, Parkville, Victoria, Australia
CHAPTER
8

Genetic Epidemiology of Ascaris: Cross-transmission between Humans and Pigs, Focal Transmission, and Effective Population Size

Charles D. Criscione
Texas A&M University, College Station, TX, USA

OUTLINE

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>203</td>
</tr>
<tr>
<td>Ascaris Cross-Transmission Between Humans and Pigs</td>
<td>205</td>
</tr>
<tr>
<td>Landscape Genetics as a Means to Infer Ascaris Transmission within a Host Population</td>
<td>208</td>
</tr>
<tr>
<td>Effective Population Size: Epidemiological Utility and Estimation</td>
<td>211</td>
</tr>
<tr>
<td>Concluding Remarks</td>
<td>224</td>
</tr>
<tr>
<td>References</td>
<td>226</td>
</tr>
</tbody>
</table>

INTRODUCTION

In many regards, the field of genetic epidemiology (a.k.a. molecular or evolutionary epidemiology and here defined as the use of genetic/molecular markers to infer some aspect of the parasite/pathogen’s
population biology such as transmission, population growth, or selected traits) asks the same questions as asked in the field of conservation genetics. Is there just one species or are there cryptic evolutionary units, is the species fragmented into subpopulations, was the fragmentation the result of human perturbation, is the population declining, what facilitates connectivity/gene flow among subpopulations, what was the source of invasion (outbreak) for an exotic species (emerging pathogen), what loci are of adaptive significance? The key difference between epidemiology and conservation is the end goal. Epidemiologists try to eliminate or reduce populations of parasites/pathogens. In contrast, conservationists strive to maintain or increase population sizes and continuity of endangered species. Population genetic applications are now integral in conservation because it is well recognized that low genetic diversity, small effective population sizes, and population fragmentation (all three of which can be measured via genetic methods) can increase the chance of population extinction. Because conservation geneticists are interested in these factors to prevent extinction, then it seems logical that epidemiologists could use similar data to help reduce or eradicate parasites/pathogens. Indeed, because of the parallel questions between the fields, much of the population genetics theory, methods, and reasoning that are used in conservation genetics could be applied to genetic epidemiology. For instance, it is recognized that low genetic diversity can reduce evolutionary potential (i.e. the ability of populations to evolve to cope with environmental change). Chemotherapy control programs are a major environmental change for parasites. Given that drug resistance has evolved among several helminths, it seems reasonable that reducing genetic diversity, via a reduction in effective population size (discussed below), should be an imperative epidemiological goal to help prevent drug-resistant evolution.

In this chapter, I discuss three pertinent applications of population genetics (all of which have been utilized in conservation biology) to further our understanding of *Ascaris* epidemiology in fine scale geographic studies. First, I focus on whether sympatric populations of *Ascaris* in humans and pigs constitute separate populations in order to ascertain if there is cross-transmission between human and pig hosts. Second, I discuss the use of landscape genetics to identify foci of transmission and epidemiologically relevant variables correlated to substructure of parasite populations. These first two topics correspond to a series of recently proposed hierarchical questions aimed at addressing local scale population genetics in metazoan parasites. Thus, I refer readers to Gorton and colleagues for a more general discussion of these topics in metazoan parasites. Also, these sections are not intended to be a comprehensive summary of the *Ascaris* population genetics literature as this was recently reviewed by Peng and Criscione. The third section
of this chapter proposes the novel integration of the effective population size \( (N_e) \) parameter into population monitoring and epidemiological studies of parasites. Using microsatellite data from a metapopulation of \( A. \ lumbricoides \) in Nepal, I demonstrate the utility of estimating \( N_e \) with single-sample, contemporary estimators. I also discuss assumptions and provide some guidelines for estimating \( N_e \). My goal is to emphasize the importance of the above topics in epidemiological research, highlight the population genetic methodologies that have been used, and point to new directions that may aid the development or monitoring of \( Ascaris \) (and metazoan parasites in general) control programs.

A species’ life history and the way samples are collected can influence interpretation of some of the genetic analyses I discuss. Thus, I first provide a brief summary of the biological characteristics of \( Ascaris \). Sampling will be addressed in the context of each study that is discussed below and just note here that genotypes were always obtained from adult worms. \( Ascaris \) has a direct life-cycle where mature male and female adult worms reside in the lumen of the small intestine.\(^8\) The mating system has not been extensively studied. However, recent paternity analyses indicate there is polyandry in pig \( Ascaris \)\(^9\) and Hardy–Weinberg equilibrium, indicating random mating, has been observed on very local scales (i.e. within people in a single village).\(^10\) A female can produce millions of eggs over her lifetime, which is about 1 year.\(^11\) Eggs are released into the external environment where they can persist for 6 to 9 years.\(^12\) Infection occurs by ingestion of eggs via fecal contaminated material. Larvae hatch in the small intestine, penetrate the intestinal wall, migrate to the lung to become fourth-stage larvae, and then migrate up the trachea back into the esophagus and ultimately the small intestine. In about 60 days from the point of infection, females will start to produce eggs.\(^8\) Key life history aspects in terms of population genetics are that breeding worms are transiently separated into groups (i.e. hosts)\(^13\) and that the long-lived eggs can lead to overlapping generations. As will be discussed, the latter is of significance because breeding worms that end up in the same host may be of different offspring cohorts (i.e. there is overlapping of generations).

**ASCARIS CROSS-TRANSMISSION BETWEEN HUMANS AND PIGS**

The subject of whether \( Ascaris \) in humans and pigs is one or two species (\( A. \ lumbricoides \) and \( A. \ suum \), respectively) is still being discussed\(^7,14,15\) and really points to an underlying question that is central for many human parasites: are there reservoir hosts (i.e. is there zoonotic transmission)? The answer to this question would clearly impact control
strategies in terms of which hosts should be targeted: just humans or both humans and pigs? It is clear from mitochondrial sequence (mtDNA) data that there is strong neutral genetic differentiation between roundworms originating from sympatric host species.\textsuperscript{16–18} These data indicate there is non-random transmission between the host species such that there is not a single source pool of infection shared by humans and pigs. However, because there were no fixed allelic genetic differences between human and pig \textit{Ascaris} samples, these results were unable to ascertain if there were two completely independent transmission cycles (one through humans and one through pigs) or if there was limited cross-infection between the two host species. The lack of fixed sequence differences could result from incomplete lineage sorting (retention of ancestral lineages in descendent taxa) with no cross-transmission, current introgression (hybrid offspring resulting from cross-breeding between human and pig \textit{Ascaris}), or cross-transmission, but no interbreeding (e.g. a worm is a first generation migrant from one host species to the other).\textsuperscript{7} In areas of non-endemic human transmission (USA, Denmark, and Japan), worms obtained from humans had DNA sequences that matched those obtained from pigs.\textsuperscript{19–21} These data clearly show cross-transmission from a pig source into humans and raise the possibility that the lack of fixed differences observed in human–pig endemic areas is also due to cross-transmission. Thus, two important questions are raised: (1) how can one detect cross-transmission in human–pig endemic sites, and (2) if there is cross-transmission, is there introgression between human and pig \textit{Ascaris}?

Criscione and colleagues\textsuperscript{22} addressed these questions with genetic-based assignment/model-based clustering methods.\textsuperscript{23,24} These methods, which have a history in species management applications, use information from multilocus genotypes (commonly assuming Hardy–Weinberg equilibrium and linkage equilibrium among loci) to ascertain population membership of individuals.\textsuperscript{25} They can also be used for identifying first generation migrants and hybrid individuals. Genetic assignment/model-based clustering methods provide several advantages for allowing one to detect hybrids. First, analyses can be conducted when no taxa-specific markers exist,\textsuperscript{26,27} as is currently the case with \textit{Ascaris} of humans and pigs.\textsuperscript{7} Second, separate samples where each only contains individuals of a single parental population are not required.\textsuperscript{26} Third, \textit{a priori} delineation of populations is not necessary (i.e. no knowledge of underlying substructure is needed for the analyses). The latter is important as the finding of cryptic species and substructure is not uncommon among metazoan parasites.\textsuperscript{6}

From both a village in Guatemala and a county in the Hainan Province of China, Bayesian clustering methods with genotypes of adult worms clearly delineated genetically structured parasite populations between
human and pig hosts in sympatry. These results were in accordance with previous mtDNA-based studies. Moreover, the multilocus genotype data enabled the identification of hybrid worms (4% in Guatemala and 7% in China). The finding of hybrids necessarily implies that there was cross-transmission between human and pig hosts because a worm of pig origin and of human origin had to meet in the same host in order to mate. This cross-transmission and interbreeding had to be recent as the methods employed can only detect hybrids going back two generations. Zhou and colleagues used the same methods to ascertain the frequency of cross-transmission across six provinces in China. They observed similar results and identified both first generation migrants (~7% of sampled worms) and hybrid worms (also ~7%), both of which were predominantly collected from human hosts. Notably, the authors state “The results strongly suggest pig ascaris as an important source of human ascariasis in endemic area where both human and pig ascaris exist. In consideration of current control measures for human ascariasis targeting only infected people, it is urgently needed to revise current control measures by adding a simultaneous treatment to infected pigs in the sympatric endemics”. With these new molecular tools at hand, it will be prudent to perform additional studies from sympatric populations to determine if limited cross-transmission is a global theme especially in relation to different pig-raising, cultural, or economic conditions. It will also be of interest to see if cross-transmission continues to show a largely pig to human pattern and to explore the mechanisms that generate the genetic differentiation between the host-associated populations despite the high frequencies of cross transmission.

Aside from the direct inference of cross-transmission, what is the epidemiological significance of limited cross-transmission and introgression? Criscione and colleagues highlight two critical aspects. First, while there is significant genetic differentiation between ascaris populations in humans and pigs, the long-term ability to cross-transmit between host species remains possible. Thus, even in non-endemic sites, human infection via a pig source remains possible (as evidenced by several studies). Also, this ability may have led to a complex evolutionary history of multiple host colonization events. Second, hybridization can lead to introgression of adaptive genes and hybridization itself may produce new combinations of parasite genotypes that increase parasite virulence or host range via host immune evasion. Little attention has been given to these aspects of ascaris epidemiology. Because parasite hybridization is of long-term epidemiological significance in terms of the evolution of novel host infectivity genes or drug-resistant genes, it will be critical to begin mapping regions of genomic introgression in relation to host species infectivity patterns in ascaris.

III. EPIDEMIOLOGY OF ASCARIASIS
Effective *Ascaris* control will require detailed knowledge of parasite dispersal to fully evaluate transmission patterns among individual human hosts. The extent of parasite dispersal, however, is difficult to ascertain with data based solely on infection intensities (i.e. number of worms per infected host or a surrogate such as eggs per gram of feces). This is because direct observation of parasite offspring leaving one host and subsequently infecting the same or a new host is nearly impossible.\(^31\) Thus, while intensity data are necessary to explore factors that explain the variation in the distribution of parasites among individual hosts,\(^32\) they do little to answer the question of where did an individual acquire their infection (i.e. are there different foci of infection in the single human population).

Identification of population subdivision via population genetics analyses of multilocus genotypic data provides a powerful means to infer macroparasite dispersal among subdivided units such as individuals or groups of hosts (e.g. households).\(^33\)–\(^36\) When using genetic data to infer transmission among individual hosts, the sampling unit should be the parasite stage that infects that host.\(^6\) In the case of *Ascaris*, adult worms would be genotyped from human hosts. If, for example, expelled *Ascaris* eggs from humans were used, then measures of genetic differentiation could be inflated due to the possible sampling of sibling parasites. I refer readers to Steinauer and colleagues\(^37\) for a more thorough discussion of this type of sampling. Additional insight into what controls the transmission process can be gained by using landscape genetic statistical approaches to test if epidemiological variables correlate with the observed parasite genetic structure. Landscape genetics is a multidisciplinary field that incorporates spatial statistics, landscape ecology, and population genetics to evaluate the role of landscape variables (e.g. altitude, ground cover) in shaping genetic differentiation among populations.\(^38\) In this regard, landscape genetics has parallel goals with the field of spatial epidemiology, which examines the correlates of spatial variation in infection intensity patterns.\(^39\) As landscape genetics is still a developing field where several methodologies are being explored, I refer readers to a special issue in *Molecular Ecology* that highlights this field in more detail.\(^40\) Here, I demonstrate the application of landscape genetics to the epidemiology of *A. lumbricoides* from an endemic population in Jiri, Nepal.\(^41\)

The goals of the study by Criscione and colleagues\(^41\) were to determine if there was more than one source pool of infection (i.e. foci of infection) and, if so, to examine epidemiological variables that may

---

**III. EPIDEMIOLOGY OF ASCARIASIS**
correlate with these foci. If there is high mixing and dispersal of parasites across the human population, then the parasites would have a panmictic population structure. Thus, people would effectively be acquiring infections from a common parasite population (i.e. a single source pool of infection). In contrast, repeated transmission that is localized at particular foci across the human population would limit parasite mixing, leading to parasite genetic differentiation within a single human population. The finding of multiple genetic clusters of parasites, therefore, is an indication that there could be multiple infection foci (see Figure 1 in Criscione and colleagues41). Adult *A. lumbricoides* were collected from 320 people across 165 households that spanned an area approximately 14 km². In addition to spatial sampling, two temporal samples (~3 years apart, so a total of 211 household-by-year samples) were taken for some regions of the village. For logistic reasons, temporal sampling was staggered for three regions of Jiri such that one group of houses was sampled in 1998 and 2001, a second group in 1999 and 2002, and a third in 2000 and 2003. As noted below, time of collection explained less than 1% of the variance in the genetic structure of the parasite population.41 A total of 1094 roundworms were genotyped at 23 autosomal microsatellite markers.10 Model-based Bayesian clustering (implemented in the program *STRUCTURE*23) was used to analyze the multilocus parasite genotypes to determine if there was underlying genetic structure among the sampled worms. Importantly, no prior spatial or temporal information was included (or needed) in this analysis.

There was strong support for local-scale genetic structuring with 13 genetic clusters of parasites identified. The results of the population clustering analyses were subsequently incorporated into a non-parametric multivariate analysis of variance42,43 to elucidate spatial, geographical, or epidemiological features associated with the partitioning of genetic variation among the sampled worms. This analysis provided a novel approach to integrating individual-based genetic assignment results with downstream statistical analyses.41 The independent variables included a nested design (household and hosts nested within household) and eight covariates: host age, host sex, host density (number of people living in the house), elevation, geographic distance among households (latitude—longitude combined), infection intensity, parasite sex, and time of collection. When variables were analyzed independently, household explained >63% of the variance in genetic structuring whereas each covariate always accounted for <15%. When the nested design was conditioned on the eight covariates (i.e. variance due to the covariates was accounted for first), the contribution of household was still high and explained >36%. In contrast, none of the eight covariates were significant after accounting for the nested design. Interestingly, time had no impact
on the underlying genetic structure even when compared pairwise between time periods for 18 households with sufficient sample sizes for testing. Furthermore, a spatial autocorrelation analysis showed that parasites between households within 540 m were more genetically similar than expected by chance alone. Genetic differentiation measured as $F_{CT}$ (hierarchical $F$-statistic of household to the total) was 0.023 and highly significant ($p < 0.0001$).

These results revealed three key insights into transmission of *A. lumbricoides* in Jiri: there were separate foci of transmission at this local scale, households and nearby houses shared genetically related parasites, and people reacquired their worms from the same source pool of infection over time. These results challenge the dogma that a single human community will correspond to a homogeneous parasite population (implicit in many classic models of parasite transmission that measure a single basic reproduction number, $R_0$). In Jiri, multiple source pools of infection need to be considered when modeling parasite transmission. Thus, when using models to evaluate control strategies in Jiri, it would be more appropriate to consider incorporating parasite populations that exist in an interconnected network, i.e. metapopulation.46

Although I emphasized how population genetics can be used to elucidate transmission patterns, I note that I do not view landscape genetics as a panacea for epidemiological goals in general, nor do I view genetics data as a replacement for infection intensity data. Rather I see the two types of data as providing different, but complementary, information about the transmission process. For example, Walker and colleagues found that in Bangladesh host age and sex explained part of the variation in worm burdens. In contrast, host age and sex were not correlated to how worm genetic variation was partitioned in Jiri, Nepal. I realize that data from the two studies are not directly comparable as they were from different locations, but the point is that both parasite intensity and genetic data are needed to fully elucidate the transmission process. Thus, in this hypothetical comparison, although gender may account for differences in worm burdens within a household (females have higher intensities possibly due to peridomiciliary behaviors that increase exposure), males and females are still getting their worms from the same source of infection. Lastly, it should be noted that the patterns in Jiri may not extrapolate to other locations as differences in human behavior, topography, and external environmental conditions could alter transmission patterns. For instance, a communal use of human feces for fertilizer may facilitate parasite dispersal thereby creating a single source pool of parasites. Thus, the assumption of a single infectious pool of parasites will need to be tested for each population of interest and as evidenced by the study in Jiri, even on very local scales.
The effective population size ($N_e$) is the size of an ideal population that has the same rate of genetic drift as the population under consideration. The “ideal” population follows the models of Wright \(^47\) and Fisher \(^48\) and, in simple terms, refers to the situation where every individual has an equal opportunity to contribute genes to the next generation. \(^49,50\) The effects of genetic drift can be measured several ways such as by the increase in inbreeding, increase in variance in allele frequency, or loss of heterozygosity over generations. Hence, there are different definitions of $N_e$: inbreeding $N_e$, variance $N_e$, and eigenvalue $N_e$, respectively. \(^51\) In closed populations of constant size, the different concepts have similar or identical values of $N_e$, but certain demographic scenarios can lead to different estimates of $N_e$ depending on which aspect of drift is being measured. \(^49--52\) My discussion will largely not make a distinction between the different $N_e$ concepts; however, the estimates I provide are more closely related to inbreeding $N_e$. Commonly, but not always, $N_e$ is smaller than the actual census population size ($N_c$) because some parents contribute many more offspring to the next generation than others.

Of what interest is parasite $N_e$ to epidemiologists? There are both long-term (evolutionary) and short-term (ecological) utilities of $N_e$. Evolutionary importance stems from the fact that $N_e$ directly determines the rate of drift where the loss of neutral genetic variation (often quantified via expected heterozygosity; $H_e$) each generation is expected to decline by a rate inversely dependent on $N_e$. \(^51\) $N_e$ is also needed to assess the relative importance of the three other evolutionary mechanisms (mutation, gene flow, and selection). For instance, equilibrium gene diversity in the infinite alleles model is determined by $N_e$ and the mutation rate ($u$) such that

$$H_e = \frac{\theta}{\theta + 1},$$

where $\theta = 4N_c u$. \(^51\) Additionally, if $N_e s << 1$ ($s = $ selection coefficient), change in allelic frequency is determined primarily by genetic drift rather than selection. \(^47\) Given these above relationships, it is clear why $N_e$ is an important parameter in conservation biology. \(^53\) Indeed, conservationists are concerned about populations with small $N_e$ because there is lower genetic variation to respond to environmental change (i.e. lower adaptive potential), the breeding of closely related individuals can reduce the fitness of an outbreeding species (inbreeding depression), and deleterious alleles can become fixed at low $N_e$. \(^1--3\) All of the latter may increase the chance for population extinction. \(^1\) Of course, the latter is the goal for epidemiologists. Consequently, from a disease management perspective, determining $N_e$ becomes of particular importance when planning control and eradication efforts.
perspective, reducing parasite $N_e$ has the long-term goal of helping to reduce parasite adaptive potential. Because drift affects loci across the genome, reducing parasite $N_e$ may help reduce standing genetic variation at any given locus that could become of adaptive significance in the face of drastic environmental changes (e.g. application of drugs or vaccines). Moreover, the parameter $N_e$ itself is necessary to help model the potential for drug resistance evolution in relation to the selective pressures induced by chemotherapy programs.

In ecological (epidemiological) terms, $N_e$ is important as it is directly determined by life history variation. Demographic factors such as fluctuating population size, non-binomial variation in reproductive success and unequal sex ratios can cause $N_e$ to deviate (likely lower) from $N_c$. Thus, knowledge of what demographic factors impact parasite $N_e$ might begin to help link the microevolutionary dynamics of parasites to transmission models that examine the reproductive potential and population growth of parasites. Admittedly, measuring demographic variables can be difficult in parasites. Thus, I believe that more immediate applications of using $N_e$ in epidemiological studies will stem from recent developments of single-sample, contemporary genetic estimators of $N_e$. In particular, the linkage disequilibrium ($LD-N_e$) and sibship assignment ($SA-N_e$) methods hold great promise to estimate $N_e$ in parasite populations. Because these methods require only the genotyping of a sample of parasites from a single time point, they will be useful in generating estimates of $N_e$ for the long-term applications noted above. Moreover, for short-term applications, recent simulations have shown that $LD-N_e$ estimates from two time points can be used to detect population bottlenecks or fragmentation of a population. Therefore, what I envision for short-term applications is the use of $N_e$ estimates as a genetic means to monitor parasite control programs. For instance, one can ask if a chemotherapy program not only reduces worm burdens ($N_c$), but also $N_e$. Does a control program reduce parasite dispersal across the treated population (i.e. cause population fragmentation)?

I am unaware of any study that has provided contemporary estimates of $N_e$ in a metazoan parasite of animals much less the application of $N_e$ estimates to monitoring a macroparasite control program. Genetic monitoring studies of parasites largely focus on levels of allelic richness ($A$) or $H_e$. While it is important to report the latter two statistics, there are disadvantages to these indices of genetic diversity. First, $A$ is subject to sample size unless rarefaction (i.e. subsampling larger samples to compare richness values among samples with different sample sizes) is used. Second, both $A$ and $H_e$ (or the DNA sequence data equivalent, $\pi$) are affected by mutation rate. This means these two measures provide somewhat redundant information, as $A$ increases so does $H_e$ (e.g. with two equally frequent alleles $H_e = 0.5$, with four $H_e = 0.75$). Being affected
by mutation also means comparisons across studies that use different loci may be inhibited as different loci (e.g. SNPs vs. microsatellites) may have different mutation rates. In contrast, changes in $N_e$ will be comparable across studies and species. Third, while $A$ and $H_e$ may provide an indication of immediate evolutionary potential, they have no predictive value for future levels of genetic diversity. As noted above, $N_e$ is a critical parameter in many evolutionary models including future $H_e$. Below I provide an example of how contemporary $N_e$ estimates can be used to further elucidate the epidemiology and population dynamics of human parasites.

The $N_e$ estimates in Table 8.1 were generated with genotype data from *A. lumbricoides* in Jiri (same data set as described for the landscape genetics study). I note that the study by Criscione and colleagues was not designed to address specific questions about $N_e$ or the effects of chemotherapy on parasite population dynamics. Worms were originally collected to examine how human genetics may play a role in parasite infection intensities. Thus, sampling is less than ideal for some of the questions I address below. Furthermore, I am assuredly violating certain assumptions for some of the population genetic theoretical models that I utilize below. I try to highlight where some of these assumptions may be violated. However, I encourage readers to research the references for the models as space limitations prevent an in-depth discussion of all assumptions. My main goal in going through several models is to show epidemiologically related questions one could ask with $N_e$ and to highlight some sampling issues associated with estimating $N_e$. Nonetheless, despite the assumptions I make, I believe the presented data do provide a reasonable approximation for some important population dynamics of *Ascaris* in Jiri.

I am primarily interested in estimating the parasite $N_e$ from households (i.e. subpopulations of the *Ascaris* metapopulation in Jiri). As discussed previously, there was significant parasite genetic structure across Jiri that was largely explained by households (>63%). As there was focal transmission around households, this would be the scale by which one would monitor the impact of a control program on parasite population dynamics. Also, because of the genetic subdivision, the $N_e$ of subpopulations will be of relevance in relation to adaptive potential (i.e. this is the level by which one would monitor genetic diversity or model the relative influence of genetic drift versus selection). In my data set, household $N_e$ was estimated from a sample of adult *Ascaris* that were collected from individual people of a household after chemotherapy treatment. A critical aspect to consider is what effective size is being estimated from this collected sample. This is outlined in Figure 8.1. Because *Ascaris* has long-lived egg stages in the external environment, the effective number of adults breeding in year $t$ ($N_t$) will have a proportion of their offspring

---

**III. EPIDEMIOLOGY OF ASCARIASIS**
<table>
<thead>
<tr>
<th>House ID_year&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Genotyped worms&lt;sup&gt;b&lt;/sup&gt;</th>
<th>House intensity&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Allele freq. cutoff&lt;sup&gt;d&lt;/sup&gt;</th>
<th>LD-&lt;i&gt;N&lt;/i&gt;&lt;sub&gt;e&lt;/sub&gt;</th>
<th>LD 95% lower</th>
<th>LD 95% upper</th>
<th>SA-&lt;i&gt;N&lt;/i&gt;&lt;sub&gt;e&lt;/sub&gt;</th>
<th>SA 95% lower</th>
<th>SA 95% upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>014_1999</td>
<td>11</td>
<td>19</td>
<td>0.05</td>
<td>18.1</td>
<td>12.9</td>
<td>27.7&lt;sup&gt;f&lt;/sup&gt;</td>
<td>110</td>
<td>37</td>
<td>infinite</td>
</tr>
<tr>
<td>014_2002</td>
<td>13</td>
<td>13</td>
<td>0.04</td>
<td>21.4</td>
<td>15.6</td>
<td>31.6</td>
<td>52</td>
<td>24</td>
<td>447</td>
</tr>
<tr>
<td>076_1999</td>
<td>10</td>
<td>29</td>
<td>0.06</td>
<td>20.5</td>
<td>13.9</td>
<td>34.5&lt;sup&gt;f&lt;/sup&gt;</td>
<td>180</td>
<td>57</td>
<td>infinite</td>
</tr>
<tr>
<td>077_2000</td>
<td>10</td>
<td>12</td>
<td>0.06</td>
<td>104</td>
<td>34.5</td>
<td>infinite</td>
<td>180</td>
<td>53</td>
<td>infinite</td>
</tr>
<tr>
<td>080_2000</td>
<td>12</td>
<td>18</td>
<td>0.05</td>
<td>–303.3&lt;sup&gt;e&lt;/sup&gt;</td>
<td>89.3</td>
<td>infinite</td>
<td>109&lt;sup&gt;e&lt;/sup&gt;</td>
<td>1</td>
<td>infinite</td>
</tr>
<tr>
<td>092_2000</td>
<td>11</td>
<td>13</td>
<td>0.05</td>
<td>24.3</td>
<td>14.8</td>
<td>52.7</td>
<td>44</td>
<td>20</td>
<td>635</td>
</tr>
<tr>
<td>097_2000</td>
<td>22</td>
<td>32</td>
<td>0.03</td>
<td>42.5</td>
<td>34.1</td>
<td>55.3</td>
<td>116</td>
<td>59</td>
<td>444</td>
</tr>
<tr>
<td>097_2003</td>
<td>37</td>
<td>43</td>
<td>0.02</td>
<td>66.7</td>
<td>54.8</td>
<td>83.9</td>
<td>133</td>
<td>79</td>
<td>272</td>
</tr>
<tr>
<td>119_2002</td>
<td>27</td>
<td>28</td>
<td>0.02</td>
<td>41.3</td>
<td>34.3</td>
<td>51</td>
<td>117</td>
<td>67</td>
<td>321</td>
</tr>
<tr>
<td>121_1999</td>
<td>29</td>
<td>89</td>
<td>0.02</td>
<td>90.1</td>
<td>67.4</td>
<td>132.3</td>
<td>180</td>
<td>100</td>
<td>862</td>
</tr>
<tr>
<td>121_2002</td>
<td>13</td>
<td>15</td>
<td>0.04</td>
<td>53.3</td>
<td>30.2</td>
<td>171.8&lt;sup&gt;f&lt;/sup&gt;</td>
<td>104</td>
<td>48</td>
<td>infinite</td>
</tr>
<tr>
<td>122_1999</td>
<td>82</td>
<td>173</td>
<td>0.02</td>
<td>314.8</td>
<td>211.5</td>
<td>582.6</td>
<td>251</td>
<td>183</td>
<td>363</td>
</tr>
<tr>
<td>122_2002</td>
<td>42</td>
<td>48</td>
<td>0.02</td>
<td>271.8</td>
<td>178.3</td>
<td>546.6</td>
<td>265</td>
<td>164</td>
<td>680</td>
</tr>
<tr>
<td>123_1999</td>
<td>33</td>
<td>115</td>
<td>0.02</td>
<td>59.2</td>
<td>48.8</td>
<td>74.1</td>
<td>92</td>
<td>58</td>
<td>170</td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>----</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>123_2002</td>
<td>21</td>
<td>31</td>
<td>0.03</td>
<td>146.7</td>
<td>78.5</td>
<td>785.3</td>
<td>420</td>
<td>131</td>
<td>infinite</td>
</tr>
<tr>
<td>124_1999</td>
<td>11</td>
<td>34</td>
<td>0.05</td>
<td>447.7</td>
<td>43.4</td>
<td>infinite</td>
<td>2.15 x 10^9</td>
<td>e</td>
<td>1</td>
</tr>
<tr>
<td>128_1999</td>
<td>13</td>
<td>24</td>
<td>0.04</td>
<td>28</td>
<td>19.7</td>
<td>45.2</td>
<td>156</td>
<td>64</td>
<td>infinite</td>
</tr>
<tr>
<td>133_2002</td>
<td>15</td>
<td>20</td>
<td>0.04</td>
<td>34.7</td>
<td>25.5</td>
<td>51.9</td>
<td>f</td>
<td>210</td>
<td>83</td>
</tr>
<tr>
<td>134_1999</td>
<td>14</td>
<td>32</td>
<td>0.04</td>
<td>54.8</td>
<td>32.6</td>
<td>142.7</td>
<td>f</td>
<td>364</td>
<td>95</td>
</tr>
<tr>
<td>135_1999</td>
<td>65</td>
<td>132</td>
<td>0.02</td>
<td>240</td>
<td>178.1</td>
<td>359.2</td>
<td>208</td>
<td>147</td>
<td>310</td>
</tr>
<tr>
<td>135_2002</td>
<td>72</td>
<td>85</td>
<td>0.02</td>
<td>184.5</td>
<td>147.4</td>
<td>242.7</td>
<td>173</td>
<td>125</td>
<td>246</td>
</tr>
<tr>
<td>140_1998</td>
<td>23</td>
<td>31</td>
<td>0.03</td>
<td>40.7</td>
<td>31.9</td>
<td>54.6</td>
<td>67</td>
<td>36</td>
<td>195</td>
</tr>
<tr>
<td>140_2002</td>
<td>14</td>
<td>20</td>
<td>0.04</td>
<td>183.2</td>
<td>57.6</td>
<td>infinite</td>
<td>364</td>
<td>113</td>
<td>infinite</td>
</tr>
<tr>
<td>148_1998</td>
<td>13</td>
<td>23</td>
<td>0.04</td>
<td>-1590.5</td>
<td>e</td>
<td>78.9</td>
<td>infinite</td>
<td>312</td>
<td>101</td>
</tr>
<tr>
<td>148_2002</td>
<td>22</td>
<td>24</td>
<td>0.03</td>
<td>89.1</td>
<td>57.2</td>
<td>185.3</td>
<td>185</td>
<td>89</td>
<td>12,788</td>
</tr>
<tr>
<td>152_1998</td>
<td>12</td>
<td>22</td>
<td>0.05</td>
<td>37.9</td>
<td>22.6</td>
<td>94.5</td>
<td>f</td>
<td>132</td>
<td>47</td>
</tr>
</tbody>
</table>

*Household identification numbers correspond to those in Figure 2 of Criscione and colleagues.*

The number of worms that were genotyped at 23 microsatellite markers per household-by-year. Raw data are from Criscione and colleagues.

The total number of worms collected per household-by-year after albendazole treatment. See Criscione and colleagues for details of sampling.

Alleles with frequencies below this value were omitted when estimating \( N_e \) with the LD method.

Negative or 2.15 x 10^9 estimates of \( N_e \) are regarded as infinite (see text for explanation).

The LD-\( N_e \) method had an upper bound for the 95% CI for the given allele frequency cutoff, but at other cutoffs, estimates typically included infinity as the upper bound. In contrast, LD-\( N_e \) estimates in shaded rows often provided bounded CI even at other allele frequency cutoffs.

The jackknife method was used for the LD interval and the SA interval is estimated in the program. Estimates were generated with the programs LDNE and COLONY, respectively. The 13 shaded rows highlight where both estimators yielded \( N_e \) estimates bounded by confidence intervals.
FIGURE 8.1  Diagram showing how the life history of *Ascaris* and sampling relate to the estimation of effective size parameters. (A) Illustration of how the long-lived egg stage leads to overlapping generations in a single subpopulation (e.g. a household in the current study). Boxes represent the effective number of adults breeding in year $t$ ($N_t$). Five breeding years are shown and an arbitrary year is chosen as year $t = 0$. Generation length ($T$; average age to adulthood) is not known for *Ascaris*. For demonstration, $T = 2$ is shown. As an example, 25% of the offspring from $N_0$ ($L_0$) will become adults in year 1, 50% in year 2, and 25% in year 3 (dashed curved arrows). In a given breeding year, adult worms will be of mixed ages (i.e. they originate from different temporal breeding cohorts). For instance, $N_4$ will be a mixture of 25% offspring from year 1 ($L_1$), 50% $L_2$, and 25% $L_3$ (solid curved arrows). The use of a single-sample, genetic estimator (e.g. LD-$N_e$ or SA-$N_e$) on a random sample of adult worms across hosts within a subpopulation (dotted oval) provides an estimate of the generational $N_e^{56}$ (see text). Generational $N_i$ is $= T \tilde{N}_t$, where $\tilde{N}_t$ is the harmonic mean of the $N_t$’s within a generation.$^{65}$ (B) Illustration of how breeders within a given year (figure is shown for $N_4$) are subdivided among individual hosts. As noted above, adult worms will be of mixed ages (e.g. $A_1$, $A_2$, $A_3$). Eggs passed from each host are the offspring of year 4 breeders ($L_4$). The use of a single-sample, genetic estimator on a random sample of eggs from single host (dotted oval) such as might be obtained from a fecal sample provides an estimate of the effective number of breeders in that host ($N_{bi}$; dotted arrow). The equation to calculate $N_t$ is shown on the right and is a function of the $N_{bi}$’s and the $X_i$’s, where $X_i$ is the proportional contribution of progeny from the $i$th host to the mixed pool that makes up the next generation of parasite breeders. Note that if the species had discrete generations, $N_t$ is $N_e$. See Criscione and Blouin$^{13}$ for more thorough discussion of using a model of subdivided breeders to estimate parasite $N_e$. 

III. EPIDEMIOLOGY OF ASCARIASIS
that survive to reproduce in years $t + 1$, $t + 2$, and so on (Figure 8.1A). These proportions determine the average age to adulthood (i.e. generation length, $T$). Thus, even though *Ascaris* adults may live only a year in their host, generation length is likely several years longer due to the fact that eggs can persist 6—9 years in the environment. Interestingly, *Ascaris* life history closely approximates that of semelparous, age-structured species such as annual plants with seed banks and Pacific salmon. A detailed theoretical treatment of estimating $N_e$ in the latter groups of organisms is given by Waples. In short, generational $N_e$ is $\approx T\bar{N}_t$, where $T$ is generation time in years and $\bar{N}_t$ is the harmonic mean of the $N_t$'s within a generation. An important point to recognize is that a sample of adult worms of a given breeding year will contain individuals of mixed ages, i.e. there are overlapping generations (Figure 8.1A). With the LD-$N_e$ and SA-$N_e$ methods, the estimated $N_e$ reflects that of the breeders that produced the sampled adult worms (i.e. the parents of the sampled worms) and not the sampled worms themselves (i.e. not $N_t$). While cautioning that testing is needed, Waples and Do conjectured that a mixed-aged sample that includes a number of consecutive age classes approximately equal to generation length should produce an estimate roughly corresponding to generational $N_e$. Thus, throughout the chapter, I will assume that the sample of adult worms from each household provides an estimate of generational $N_e$ of each subpopulation (Figure 8.1A). I will return to the estimation of $N_t$ (Figure 8.1B) in my concluding remarks.

I used two single-sample, contemporary estimators, LD-$N_e$ and SA-$N_e$, as implemented in the programs LDNE and COLONY v2.0.2.1, respectively. Both of these methods provide estimates that are related to the inbreeding $N_e$. The LD-$N_e$ method can be sensitive to rare alleles, thus I followed the recommendations of Waples and Do for using alleles with frequencies above a cutoff given the sample size (see Table 8.1). The random mating system option was used. In COLONY, I selected the male and female polygamy options without inbreeding. These latter options in the two programs seem reasonable given the current state of knowledge about *Ascaris* mating systems. Length of run and likelihood precision (full-likelihood) were set to medium in COLONY. I used the update allele frequency option and the complexity prior, which should result in a higher $N_e$ estimate (compared to not using it) as this prior discourages complex pedigree inference.

Table 8.1 provides the estimates of $N_e$ per household-by-year where 10 or more worms were genotyped ($n = 26$). There are several important patterns and questions that emerge from these data. First, sample size matters in obtaining estimates that are not infinite or do not have an upper confidence interval of infinity. Infinity estimates (negative values in the LD-$N_e$ method or the $2.15 \times 10^9$ values in the SA-$N_e$ method) result when
sampling error swamps the genetic signature of genetic drift in the case of LD-\(N_e\) estimates\(^{56}\) or when little to no pedigree structure is found in the SA-\(N_e\) method. Of the 26 estimates, only 13 (Table 8.1, shaded rows) gave values that had bounded confidence intervals for both estimators. When looking at the other 13 estimates, it appears that several of the LD-\(N_e\) estimates had upper confidence limits when the SA-\(N_e\) method did not. However, it is important to note that these LD-\(N_e\) estimates (white rows and marked in Table 8.1) were sensitive for the allele frequency cutoff such that other cutoff values returned an infinity upper bound (data not shown). In contrast, LD-\(N_e\) estimates in the shaded rows had upper bound confidence intervals regardless of allele frequency cutoff. Thus, there was congruence between the two methods in returning estimates with uncertainty in the upper confidence limits for the same household-by-year samples. Thirteen of the 13 estimates with uncertainty in the upper confidence limits (white rows) had \(n \leq 21\), whereas 11 of the 13 estimates with bounded confidence intervals had an \(n > 21\) (Table 8.1). Small sample sizes will only provide bounded confidence intervals if the true \(N_e\) is small (\(<50\)), which is likely the case for houses 014_2002 and 092_2000 (Table 8.1). The reason is that the larger the true \(N_e\) and the smaller the sample size, the less likely one is to find related individuals in the sample (Table 2 in Waples and Waples\(^{68}\)). Thus, if small sample sizes yield estimates with unbounded confidence limits it is difficult to ascertain whether the true \(N_e\) is large or whether it is small, but a larger sampling error is to blame. If one wants to detect populations that have a true \(N_e\) of 500–1000, sample sizes need to be around 50 with about 20 polymorphic loci.\(^{56}\) It appears my current data set was able to get bounded confidence limits with \(n = 22–40\) because true \(N_e\) of each subpopulation was likely much less than 500. Several studies\(^{56–59}\) have used simulations to address sampling, thus I refer readers to these papers for a discussion of appropriate samples sizes and number of loci to use in relation to types of questions one might ask with \(N_e\) estimates.

Interestingly, almost all point estimates range in the mid tens to low hundreds. Even the unbounded confidence interval estimates, which still can give some indication of the lower bounds of \(N_e\), tend to show low \(N_e\) point estimates. From here on, however, I will restrict my analyses and discussion to the 11 estimates that had \(n > 21\) (Table 8.2). Even though houses 014_2002 (\(n = 13\)) and 092_2000 (\(n = 11\)) had estimates with bounded confidence intervals, I removed them from subsequent analyses to avoid bias. Bias may originate because I would be omitting the other houses with \(n \leq 21\) that potentially really do have larger effective sizes, but could not get an accurate estimate due to small sample size. There was a high correlation between the point estimates of the two estimators (\(r = 0.894, p = 0.0002, n = 11\); Table 8.2). These data show good congruence between the two estimators and give me high confidence I am getting
accurate estimates of the parental breeding population $N_e$ that contributed to the infections in each household. This is especially true given the two methods utilize very different methods (linkage disequilibrium versus identification of pedigree structure) to estimate $N_e$. The harmonic means of the household $N_e$ point estimates ($n = 11$) were 76.9 (95% CI: 55.6–116.8) and 137.9 (95% CI: 108.2–183.6) for the LD-$N_e$ and SA-$N_e$, respectively (CI based on 1000 bootstraps over the point estimates of the household-by-year samples). The harmonic mean is used because the distribution of $N_e$ can be highly skewed. Waples and Do also suggest that if two single-sample estimators are independent and are estimating the same parameter from a population, then a more precise or “best” estimate of $N_e$ can be obtained by taking the harmonic mean of the two single-sample estimators. The “best” estimate-$N_e$’s for the 11 household-by-year samples with $n > 21$ are given in Table 8.2. The harmonic mean of these “best” estimates is 98.8 (bootstrap 95% CI: 73.5–139.1).

One of the questions that can be asked with these data is whether or not drug treatment impacted $N_e$. Simulations have shown that LD-$N_e$ estimates from two time points can be used to detect a population genetic bottleneck. After omitting samples with small $n$, I only had three houses (97, 122, 135) with estimates from both time periods (people in households were treated and worms collected, then three years later this was repeated). This is a small sample size, but visual (i.e. not statistical)
assessment of the values and their confidence intervals (Table 8.1) does not reveal any discernible impact of chemotherapy on the $N_e$ of these *Ascaris* subpopulations (even if houses with small $n$ are examined). These genetic results parallel prior epidemiological data from Jiri where after 1 year of treatment both prevalence (year 1 = 27.2%, year 2 = 24.2%) and mean number of worms expelled per individual (2.37 and 2.67) showed little change.\(^6^2\)

The latter scenario also begs the question of whether $N_e$ reflects $N_c$ of *Ascaris* in Jiri. Intuitively, as $N_c$ increases, so should $N_e$. However, I caution that the relationship between $N_e$ and $N_c$ under different demographic scenarios is generally not well understood and may vary considerably among species.\(^6^9\) In some free-living organisms, the ratio of $N_e/N_c$ decreases as population density increases.\(^6^9\) Experimental data in flour beetles suggest this may be caused by an increase in the variation in reproductive success among individuals as $N_c$ increases.\(^7^0\) Therefore, it might be that there is an asymptotic relationship between $N_e$ and $N_c$ such that $N_e$ levels off even as $N_c$ gets larger (Figure 8.2). The latter relationship would be important in epidemiological studies because a drop in $N_c$ may not constitute a drop in $N_e$ until a critical $N_c$ is reached. This would be crucial in terms of the evolution of drug resistance because a huge selection pressure via chemotherapy could be imposed on the population without a drop in $N_e$. Selection is more efficient with larger $N_e$. Thus, both worm count and genetic data are warranted in epidemiological studies if

![FIGURE 8.2 Hypothetical asymptotic relationship between $N_e$ and $N_c$. Dotted line denotes critical $N_c$ where $N_e$ no longer substantially increases as $N_c$ increases. One possible explanation for this pattern is that as population density increases, variation in reproductive success may increase considerably, thus substantially reducing $N_e$. An important epidemiological implication is that above this critical point, $N_c$ could be reduced drastically without a dramatic effect on $N_e$. In relation to the *Ascaris* data presented, the correlation between household intensities and the single-sample $N_e$ estimates may suggest *Ascaris* subpopulations already exist below the critical $N_c$ value.](image-url)
one of the goals is also to reduce genetic diversity/adaptive potential. I did not have a means to estimate the $N_c$ of the parents that produced the sampled worms from each household especially since the parents are likely from different breeding years (i.e. different $N_t$'s; Figure 8.1A). As a surrogate, I tested for a correlation between the “best” estimate-$N_e$’s and the infection intensities recorded for each house (Table 8.2), which implicitly assumes large census populations beget large populations. There was a significant correlation between infection intensity and “best” estimate-$N_e$ ($r = 0.61$, $p = 0.047$, $n = 11$). Thus, in these samples $N_e$ may be a good tracker of $N_c$. It would be encouraging if this result holds in future studies because that means $N_e$ estimates might be useful to monitor not only adaptive potential, but also intensity data following an Ascaris treatment program. More data are certainly needed, but it is interesting to speculate that these correlations suggest that the Ascaris subpopulations do not exist at high densities (e.g. mean intensity per person was ~2.5 in Jiri62) where an asymptotic relationship between $N_e$ and $N_c$ would be relevant (Figure 8.2). In comparison, an asymptotic relationship may be more pertinent in parasites that have high infection intensities per host (hundreds to thousands) such as several trichostrongylid nematodes of livestock. Interestingly, among nematodes, the latter group is largely where drug resistance has been reported.4,5

The overall metapopulation $N_e$ ($N_{eT}$) is also of interest in relation to dynamics that occur among subpopulations (e.g. equal subpopulation contributions to the migrant pool versus extinction/recolonization dynamics). My goal in this section is to compare an estimate of $N_{eT}$ using Wright’s island model71 to an estimate of $N_{eT}$ from the single-sample estimators. I caution the combining of samples across subpopulations (and across years as in this data set) and the subsequent use of these single-sample estimators has not been quantitatively tested as a means to estimate $N_{eT}$. Thus, the following should be treated as a thought exercise rather than definitive conclusions. I used the entire data set of 1094 worms and obtained an LD-$N_e$ estimate of 1062 (95% CI: 975–1161, at the 0.02 cutoff) and SA-$N_e$ estimate of 1645 (95% CI: 1502–1789). The harmonic mean of these two estimates yields a “best” estimate of $N_{eT} = 1291$. In Wright’s island model, $N_{eT}$ is a function of subpopulation $N_e$ and genetic differentiation ($F_{ST}$) such that

$$N_{eT} \approx \frac{nN_e}{1 - F_{ST}}, \tag{8.2}$$

where $n$ is the number of subpopulations and each subpopulation has the same $N_e$. This model assumes that subpopulations contribute equally to the migrant pool. As can be seen in Eq. (8.2), as genetic differentiation increases among subpopulations, $N_{eT}$ can exceed the sum of the subpopulation effective sizes.49,72 This is because while each subpopulation

III. EPIDEMIOLOGY OF ASCARIASIS
loses variation due to drift, each subpopulation will become fixed for
different alleles. Thus, genetic variation is maintained over the entire
metapopulation. However, in metapopulation models where some
subpopulations have greater contributions to the migrant pool than others
or where subpopulations go extinct and are recolonized via founders of
another subpopulation, $N_{eT}$ can be greatly reduced below the sum of
subpopulation effective sizes. If estimates of the three parameters
in Eq. (8.2) can be obtained to estimate $N_{eT}$, then the island model value
can be compared to the single-sample $N_{eT}$ “best” estimate to draw on
conclusions about subpopulation contributions to the migrant pool.
Criscione and colleagues reported that genetic differentiation among
households was 0.023 (the equivalent of $F_{ST}$). Furthermore, using
a Bayesian clustering method, they identified 13 core clusters, which I
will use as $n$. Obviously $N_e$ was not the same across households, but for
the purpose of illustration I will assume they were and use the harmonic
mean (Table 8.2), 98.8 (95% CI: 73.5–139.1). Based on the latter values, the
island model $N_{eT}$ is 1314 (possible range from 979 to 1851), which is in
agreement with the single-sample “best” estimate of 1291. Therefore, this
comparison suggests that *Ascaris* subpopulations in Jiri reflect more of
Wright island model rather than a metapopulation where subpopulations
have large unequal contributions to the migrant pool or recolonization–
extinction dynamics. If the latter were true, then it seems like the single-
sample estimators would be producing an estimate well below that
predicted from the island model. Readers are encouraged to delve into the
references above to get an understanding of all model assump-
tions. Here I point out two concerns in this data set. First of which is the
number of subpopulations I used in Eq. (8.2). If the landscape genetics
study did not sample all possible subpopulations, then 13, and thus the
estimate of $N_{eT}$ from the island model, would be an underestimate.
Second, I also assumed that the harmonic mean $N_e$ of the households
reflects the central tendency of the $N_e$ of the 13 genetic clusters. This seems
reasonable as households were largely composed of individuals
belonging to a single cluster. However, all clusters are not represented by
the houses in Table 8.2, and a few houses may represent the same cluster
(i.e. there is pseudoreplication).
I did not have a means to estimate $N_e$ for each subpopulation.
However, if I assume stable human population growth and infection
patterns are constant over time, I can estimate a census size for *Ascaris*
across the Jiri metapopulation ($N_{cT}$). This enables me to get a $N_{eT}/N_{cT}$
ratio. Using the average prevalence of 25.7% and intensity of 2.52 worms
per infected host data from Williams-Blangero and colleagues and the
1991 census count of the Jiri human population of 7138, the $N_{cT}$ of
*Ascaris lumbricoides* would be 4623. Accordingly, $N_{eT}/N_{cT} = 0.28$ when using
the single-sample “best” estimator for $N_{eT}$. The single-sample estimators

III. EPIDEMIOLOGY OF ASCARIASIS
used here would reflect uneven sex ratios and variation in reproductive success of the previous breeding generation. In an extensive review by Frankham, the mean $N_e/N_c$ ratio was 0.35 (95% CI: 0.28–0.42) among species for which variation in reproductive success and uneven sex ratios were taken into account to obtain demographic estimates of $N_e$. Thus, the *Ascaris* value falls just on the edge for what is known from single generation $N_e/N_c$ estimates of other species.

The following may be a bit of an extrapolation because of the restrictive assumptions of the island model, but I think it is a useful exercise in what genetic data and a $N_e/N_c$ ratio might be able to tell us. Under the assumptions of Wright’s island model genetic differentiation is a function of subpopulation $N_e$ and migration rate ($m$) where

$$FST \approx \frac{1}{4N_em + 1}. \quad (8.3)$$

As discussed above, the island model might approximate the *Ascaris* population dynamics in Jiri. Thus, it seemed reasonable to estimate the effective number of migrants per generation ($N_em$) from Eq. (8.3). Using a $FST$ of 0.02341, $N_em = 10.61$. If the $N_{et}/N_{ct}$ ratio of 0.28 also represents the ratio within subpopulations, then that means about 38 census worms per generation are migrants into the foci of transmission around households. This does not mean all 38 census worms become adults or even infect a person. It would be more appropriate to say a minimum of 10 migrant census worms infect people (necessarily adult worm infections because $N_em$ represents individuals that contribute to the gene pool), but up to 38 census worms infecting a household were acquired from another transmission focus per worm generation. A key point here is “per worm generation.” *Ascaris* adult worms live about 1 year in their host. Thus, one might conclude generation time is 1 year and, therefore, 10–38 migrant worms per year cause infections. However, as noted above, the long-lived egg stages of *Ascaris* will increase generation time. Thus, these 10–38 migrant worms will be spread out likely over several years.

Above I have focused on using single-sample estimators to estimate the $N_e$ of the parents that generated the infections in the sampled households. One can also estimate long-term or coalescent $N_e$ that reflects the historical evolutionary dynamics of a population. Such an estimate may provide a historical baseline for what the parasite’s $N_e$ was like prior to the implementation of a control program. Waples provides a summary about estimating long-term $N_e$. Here, I illustrate estimation of long-term $N_e$ with the Jiri *Ascaris* data while also highlighting some of the caveats discussed by Waples. Long-term $N_e$ requires an estimate of $\theta = 4N_eu$, which means an estimate of $u$ is also needed. Importantly, an accurate estimate of $N_e$ via an estimate of $\theta$ will be dependent on a reliable estimate of $u$; a $10\times$ change in $u$ leads a $10\times$ change in the $N_e$. 

III. EPIDEMIOLOGY OF ASCARIASIS
Model-based genealogical simulations are preferable to estimate $\theta$, though these are computationally intensive. For simplicity, I estimated $\theta$ with Eq. (8.1), which has the assumption that the population under consideration is closed to immigration. In comparison to samples from China and Guatemala, *Ascaris* from Jiri are highly genetically differentiated. Thus, on a global scale the Jiri metapopulation of *A. lumbricoides* is likely relatively isolated. Nonetheless, sampling of locations around Jiri is needed to ascertain potential regional influences on the long-term $N_e$ estimate provided below. To estimate the coalescent $N_e$ of the metapopulation, I used $H_e = 0.71$, which was reported over all 1094 genotyped nematodes; thus, $\theta = 2.45$. There are no estimates of $u$ for microsatellites in *Ascaris*; therefore, I used estimates from the nematode *Caenorhabditis elegans*. Repeat motif and length can affect $u$ so I calculated the average $u$ from the six di- and five tetra-nucleotide motif loci with lengths less than 70 repeats (mean $u = 0.000542$ and $0.0000362$, respectively) as this would reflect the microsatellite loci in my data set. I had 19 di- and 4-tetra microsatellites, and used a weighted average to obtain an estimate of $u = 0.000454$. Using this value of $u$, the coalescent $N_e = 1347$. This long-term estimate is nearly identical to the single-sample “best” estimate of $N_{eT} (1291)$.

**CONCLUDING REMARKS**

Above I discussed how population genetics data can be used to identify cross-transmission and focal transmission. In addition, I introduced $N_e$ as a means to help genetically monitor epidemiologically relevant parasites. All the methods I have used come with assumptions and require appropriate sampling. With regards to cross-transmission and focal transmission, more discussion can be found in prior studies. Here, I will conclude with a discussion of using $N_e$ estimators for parasites especially in relation to *Ascaris* biology.

Single-sample, contemporary estimators assume closed populations with discrete generations. In regards to the assumption of a closed population, simulations showed that the LD-$N_e$ estimator is little affected by migration unless $m > 0.1$, in which case an estimate from a subpopulation will approach $N_{eT}$. The latter does not appear to be an issue in this *Ascaris* data set. Because *Ascaris* has a “seed bank” life history, it clearly does not have discrete generations. When dealing with a species with overlapping generations, generational $N_e$ is of most significance for monitoring adaptive potential or modeling the effects of selection. How then can one estimate generational $N_e$ for *Ascaris*? As conjectured and assumed in this chapter, the use of single-sample estimators on a sample with a mixed-age cohort (adult worms in the case of *Ascaris*) may actually
provide an estimate of generational \( N_e \) (Figure 8.1A).\(^56\) If this holds true (currently being tested by R. Waples, personal communication), one should aim for larger sample sizes than the current data set (e.g. \( \geq 50 \) per subpopulation) in order to make sure that all potential cohorts making up a generation are sampled. If this does not hold true, extensive data collection will be needed to obtain an estimate of generational \( N_e \) (i.e. using the formula \( N_e \approx T\tilde{N}i \))\(^65\) as one will need estimates of \( T \) and the \( N_i 's \). An estimate of \( T \) for Ascaris will likely require experiments in pigs by either monitoring infections from a cohort of eggs over years or using different aged pastures (i.e. eggs left standing 1 year, 2 years, etc.) to estimate infection efficiencies of different egg ages. For now, it must suffice to say that \( T \) is likely <6–9 years as this is the current knowledge of egg longevity in the environment.\(^12\) Because parasite breeders within a given breeding year (\( N_t \)) are separated among hosts, \( N_t \) is function of the effective number of breeders within each host (\( N_b \)) of the subpopulation and the proportional offspring contributions of each \( N_b \) (Figure 8.1B).\(^13\) I refer readers to Criscione and Blouin\(^13\) for a detailed description of a model for subdivided parasite breeders that can be used to estimate \( N_t \) from measures of the \( N_b 's \). Here I draw attention to the fact that the single-sample estimators can be used to estimate the \( N_b \) of a given host. One simply would collect and genotype eggs/larvae from an individual person. Moreover, the \( N_b \) values themselves may be of epidemiological use especially if one does not have a means to directly count adult parasites in a person (e.g. schistosome parasites).\(^13\) If \( N_b \) estimates correlate to actual intensities of infection (a relationship that still warrants testing), then \( N_b \) estimates could provide a more accurate depiction of infection intensities among hosts compared to other surrogate methods such as eggs per gram of feces. \( N_b \) estimates could also be important in helping determine the role an individual host has in contributing to a parasite’s subpopulation \( N_e \) (or \( N_t \) in the case of Ascaris).\(^13\)

In this chapter, I have illustrated the feasibility of using of single-sample estimators in estimating generational \( N_e \) estimates for subpopulations (households) of A. lumbricoides. In order to illustrate different concepts and applications that one could use with \( N_e \) estimates, I have made several assumptions and extrapolations with these data. Nonetheless, these estimates have shed additional light on the population and thus epidemiological dynamics of Ascaris in Jiri. Overall, the household \( N_e \) estimates were low (~100) and it appears that they were stable over time even with chemotherapy treatment (though a more formal test is needed). Comparison of metapopulation \( N_e \) (\( N_{eT} \)) between the island model and the single-sample estimators further elucidated transmission patterns in that subpopulations appear to be contributing fairly equally to overall dispersal of Ascaris across the metapopulation. Thus, among-subpopulation dynamics were relatively stable such that this
comparison did not support extinction/recolonization dynamics. Because seed banks can slow the rate of genetic change, it may be that the long-lived Ascaris eggs (a.k.a. parasite seed bank) are what contributed to the stability in $N_e$ over time, the lack of genetic differentiation between time periods for a given household, and prevention of subpopulation extinction.

Lastly, most of my discussion has focused on the short-term inference of $N_e$, which will be comparable across studies and species, as a means to monitor the impact of control programs on genetic diversity and population dynamics. $N_e$ also provides long-term inference in relation to adaptive potential. For instance, a threefold reduction in $N_e$ from $10^4$ to $10^1$ is likely to reduce adaptive potential. However, it is appreciated that drift will be mostly irrelevant in reducing adaptive potential if the threefold reduction is from $10^7$ to $10^4$. Also, there is likely no magic $N_e$ below which all parasite species are likely to go extinct and additional demographic factors that may vary among parasitic species will also be important. Clearly what is needed are more estimates of $N_e$ from parasites before one can begin to conclude about the adaptive potential. For instance, if the small effective sizes in Jiri are reflective of Ascaris in other places, then it is interesting to speculate that the reason why drug resistance has not been reported for Ascaris is that the low effective sizes have been an impediment to the evolution of drug resistance. Indeed, even the $N_{eT}$ and coalescent $N_e$ were both low (~1300). In contrast, coalescent $N_e$ on the order of $10^6$–$10^7$ has been estimated for populations of trichostrongylid nematodes, a group with several species that have evolved drug resistance. The use of the single-sample estimators will facilitate $N_e$ comparisons among parasite species/populations that differ in life history/demographic attributes, thus allowing future studies to explore the relationship between parasite $N_e$ and adaptive potential.

References

REFERENCES

III. EPIDEMIOLOGY OF ASCARIASIS

37. Steinauer ML, Blouin MS, Criscione CD. Applying evolutionary genetics to schistosome epidemiology. *Infect Genet Evol* 2010;10:433–43.
REFERENCES


III. EPIDEMIOLOGY OF ASCARIASIS
