Introgression of trypanotolerance genes

Promotor: **Dr. Ir. J. A. M. van Arendonk**
Persoonlijk hoogleraar bij de leerstoelgroep
Fokkerij en Genetica
Wageningen Universiteit

Co-promotoren: **Dr. Ir. H. Bovenhuis**
Universitair docent bij de leerstoelgroep
Fokkerij en Genetica
Wageningen Universiteit

Dr. F. Iraqi
Scientist, International Livestock Research
Institute, Nairobi-Kenya

Introgression of trypanotolerance genes

O. Delphin Koudandé

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit
Dr. Ir. L. Speelman,
in het openbaar te verdedigen
op woensdag 13 december 2000
des namiddags te vier uur in de Aula

Introgression of trypanotolerance genes

Koudandé, O. Delphin

Doctoral dissertation, Animal Breeding and Genetics Group, Department of Animal Sciences, Wageningen University, P.O. Box 338, 6700 AH, Wageningen, The Netherlands

ISBN: 90-5808-326-8

Printed by Universal Press, Veenendaal, The Netherlands
Cover design: Wim Valen and Delphin Koudandé

Financial support for the printing of this thesis was obtained from the Dr Judith Zwartz Foundation, Wageningen, The Netherlands

Abstract

Trypanosomosis is the most widespread parasitic disease of great importance in Africa affecting human and animals. There are breeds of cattle that are trypanotolerant. Trypanotolerance is an ability of these breeds to withstand the effects of trypanosome infections. The disease impedes production in Sub-Saharan Africa. In this thesis, the main focus is on marker-assisted breeding to improve trypanotolerance and mouse is used as model organism for cattle. We have concentrated on the opportunities to exploit individual genes (QTL) that affect trypanotolerance. We have shown how the size of a breeding experiment increases with the number and the size of the QTL region to introgress. Biological factors to be considered when estimating the required size of an introgression experiment are also highlighted. Strategies to optimize designs for introgression show that selecting two lines each carrying two of the donor's QTL allele during the backcross (when introgressing three QTL) is more advantageous than monitoring simultaneously the three QTL in terms of number of animals to maintain, number of genotyping and costs involved. An experiment in mice shows that QTL introgression is feasible and successful, and that the background genotype is an important factor to be considered when analyzing the results of such an experiment. This experiment is unique. The general discussion focuses on the introgression of trypanotolerance genes in cattle with an emphasis on Sub-Saharan zones of Africa. The introduction of trypanotolerant cattle in the humid and sub-humid zones of Africa, however, should be done with care to avoid damage to the environment and bio-diversity.

Stelligen

1. Prior estimation of the required number of animals for a marker-assisted introgression program is essential before the program is implemented (*this thesis*).

2. Introgression of trypanotolerance genes increases the resistance of animals to trypanosome infection (*this thesis*).

3. Mice are excellent model organisms to validate theories developed in animal breeding.

4. "If we knew what it was we were doing, it would not be called research, would it?" Albert Einstein.

5. Marker-assisted breeding as an application of biotechnology can contribute to sustainable food production and assure food security in developing countries.

6. Biotechnology is far too expensive for individual African countries and they should, therefore, make joint efforts through international institutions (e.g. ILRI[1], IITA[2]).

7. It is amazing that inhabitants of developed countries are keen on animal welfare and ready to pay for it, while people are still fighting against hunger in developing countries.

8. If Dutch people could get rid of their diaries for one day a year, their lives would become more spontaneous.

O. Delphin Koudandé: **Introgression of trypanotolerance genes.** Wageningen University, December 13th 2000.

[1] *International Livestock Research Institute*
[2] *International Institute of Tropical Agriculture*

Dedicated to P. Laetitia, Marlène and Cédric.

Preface

"Ask, and it will be given to you; search, and you will find; knock, and the door will be opened to you." Matthew 7, 7. Thanks to God for fulfilling me with His immense grace.

I am grateful to Professors. Pim Brascamp and Johan van Arendonk for giving me the opportunity to complete a PhD study in the animal breeding and genetics group (ABGG). Thanks to Johan van Arendonk for leading me through this research and providing the right words and moral comfort whenever they were needed.

I would like to thank Henk Udo who gave me useful advice at the time the PhD project was proposed to me. Later he became a member of my PhD supervising committee making prompt and sharp comments on the different chapters of this thesis I am very grateful to him.

I am indebted to Dominique Hounkonnou who played a key role in the leave my institute granted to me to start this PhD project. I can never be thankful enough.

I would like to acknowledge Alan Teale and Fuad Iraqi who guided my first steps at the International Livestock Research Institute (ILRI) and offered me a good environment to accomplish my experiment and complete this thesis. I would like also to thank ILRI staff and particularly people from Lab.7 and from the small animal unit.

I owe many thanks to Henk Bovenhuis who did not mind to answer my various questions while I was at ILRI in Nairobi till I returned to Wageningen in January 1999. From then he consented to become my co-supervisor. I hope that after "I kicked some open doors in 1996," we finally end up with this booklet called a thesis. I appreciate your criticisms and advice.

I highly appreciate the care provided by Adrie te Pas, Ton Derik, Harm and Sybe during the five-and-half years I have been away from my family; you have been my family here, many thanks to you.

All my gratitude to Peter Thomson, Luc L. Janss, Barbara Harlizius-Schwenger and Richard Spelman who have been helpful by providing me with some of their scientific knowledge.

I have enjoyed the companion of my colleagues from ABGG, namely Chris Schrooten, Bernado Vargas, Liesbeth van der Waaij, Kosgey Isaac, Margaret Okomo, Antonia Samore, and particularly my office mates: Piter Bijma, Jascha Leenhouwers, Abe Huisman, Dirk-Jan de Koning, Jan Thijs van Kaam and the latest and lively Raffaella Finocchiaro.

Special thanks to all members of ABGG, in particular Ada Wiggerman, Maria Lippelt, Alex Hulzebosch, Bart Ducro, Pieter de Groot, Henk Vos and Marilyne Faivre.

I would also like to acknowledge the company of my colleagues from Lab.7 at ILRI-Nairobi, namely Anne Mugai, Simon Kanga and Sebastian Chenyambuga.

Many thanks to all friends I made here in Wageningen and the various colleagues and country-mates I have met at this place and with whom I could get rid of homesickness., I was very pleased with you all.

I would like to sincerely thank my friends from the Lions Club Wageningen for the warm welcome they provided to me. I have also enjoyed being with the volleyball club companions.

A special thank to Tashi Samdup who convinced me early in 1996 to join the second term courses of animal breeding that resulted later in obtaining this PhD project.

The Netherlands Foundation for the Advancement of Tropical Research (WOTRO) and ILRI are acknowledged for the financial support they have provided for this PhD project.

I would like to acknowledge Wageningen Institute of Animal Sciences and Wageningen University for providing me the environment to complete a PhD.

Finally I acknowledge the leave provided by the National Agricultural Research Institute of Benin (INRAB) to undertake this PhD project.

Contents

CHAPTER I

General introduction

Outline

Trypanosomosis is recognized as the most widespread disease of great importance in Africa. This chapter starts with a general overview of trypanosomosis in Africa and its economic impact. Subsequently, the risk represented by the current uncontrolled crossbreeding to the indigenous resource of cattle is addressed. The evidence of genetic background for trypanotolerance and what current research programs are being undertaken are reviewed in the later sections. After a brief definition of marker-assisted introgression and its principles, the origin of the mouse model used in this thesis is introduced. This chapter ends with presenting the aims and layout of the thesis.

Trypanosomosis in Africa

Trypanosomosis is a parasitic disease affecting humans (sleeping sickness) and animals. In domesticated livestock in Africa, different trypanosomes species are simultaneously involved in the disease that shows large variation in severity and duration. This variation is associated with both the host species and the trypanosome species. Three trypanosome species are important in cattle: *Trypanosoma vivax*, *T. congolense* and *T. brucei*.

African trypanosomes are flagellated protozoa living in the blood stream of infected animals, but some species (e.g. *T. brucei*) can be found in tissues and organs (heart, skeletal muscles, brain) where they cause necrosis. Trypanosomosis is a vector borne-disease and tsetse flies are the major vectors in which trypanosomes accomplish part of their life cycle. Some other biting insects and Tabanids contribute also to the transmission of trypanosomes.

Trypanosomosis affects domesticated animals (e.g., cattle, sheep, goats, pigs, dogs, cats, horses, donkeys and camels) as well as humans, mice and wild animals. Wild animals represent the reservoir for the parasites and make eradication programs difficult.

The geographic distribution of the disease in Africa usually overlaps the distribution of tsetse flies between latitudes 14°N and 29°S corresponding to the humid and sub-humid zones of the continent (Fig. 1.1). In total, 37% of the continent, covering 11 millions square km and involving 37 countries, is infested with tsetse flies (Murray and Trail 1984). MacLennan (1980) estimated that 65% of this area (approximately 7 millions square km) could be used for livestock or mixed farming without stress to the environment if trypanosomosis was controlled.

Clinically the disease is characterized by intermittent fever, increasing degree of anemia, weight loss, abortion, alterations in ovarian cycle leading to infertility and, if left untreated, may even result in the death of the infected animals (Brown et al. 1990; Paling and Dwinger 1993). As listed above, theses symptoms denote a chronic development of

Fig.1.1 Tsetse distribution Map (Adapted from Lessard *et al.*1990)

the disease that is usually observed, however, acute and lethal disease can be observed in various hosts (cattle, horse, dog) leading to death after 1-3 weeks.

No new drugs have been developed during the last 40 years and due to the increasing resistance developed by trypanosomes, some of the available drugs have been withdrawn from usage reducing the scope of choice. Chemoprophylaxis is not advisable in endemic zones where animals are under permanent pressure of infections; curative drugs are, however, more appropriate.

Prophylactic measures such as chemical spreading (insecticide) and habitat destruction aiming at eradication of tsetse flies are not environmentally sound. Trapping, use of sterile males and insecticide-impregnated screens are more sustainable methods and contribute to reduce the vector pressure surrounding herds and therefore the infection risk. Vaccination against the disease is still not possible because trypanosomes have the ability to change antigen during the infection (Murray et al. 1984; Paling and Dwinger 1993). Other opportunistic infections like helminthosis and tick-borne diseases should be also considered in the prophylaxis program. The use of indigenous cattle breeds (N'Dama,

Baoulé, Lagune and Muturu), which are more adapted to local conditions, is often proposed as an alternative to prophylactic treatment.

Limitations of control methods and limited medium-term prospects for the development of vaccine are stimuli for using trypanotolerant breeds more efficiently to increase livestock productivity in the humid, sub-humid and semi-arid regions of Africa (Murray et al. 1990).

Economic impact

Trypanosomosis is beyond doubt the single and most important constraint to livestock development in the sub-humid and non-forested portions of the humid zone of Africa. Tsetse-transmitted trypanosomosis has both direct and indirect economic impacts to livestock production.

Direct impacts are those related to condition loss of animals, reduction in milk and meat production, morbidity, infertility, mortality and control costs of the disease. According to Kristjanson et al. (1999), the disease costs approximately US $1340 million per year for livestock producers and consumers in Africa. This cost excludes losses due to reduction of manure availability and inability to use animals for draught power.

Indirect impacts are related to the opportunity cost of land and other resources currently not used for livestock production due to the presence of tsetse flies. Other indirect impacts are the reduction in crop production due to inability of diseased animals to be used for traction and the reduction in manure production. Kristjanson et al. (1999) estimated that the potential economic benefits from trypanosomosis control in terms of meat and milk surplus could amount US $702 million per year.

Some other losses are difficult to estimate. Domestic farm animals serve as a reserve readily convertible into cash to protect small farmers against an unpredictable climate and unstable commodity prices. These animals provide an outlet for damaged grains, root crops, and other crops that are not marketable or needed for human consumption. In addition, ruminants can utilize ligno-cellulosic biomass, which includes crop residues and by-products and which has no value other than adding to soil organic matter. These nutrients and the value of these products would largely be lost if they were not consumed by livestock (Savadogo 2000). Livestock serves, therefore, to transform feed into food and marketable products, adding value to farming activities, increasing income, and enhancing the biophysical and economic viability of agriculture. Also human capital is threatened by some species of trypanosome (*T. gambiense* and *T. rhodesiense*) responsible for sleeping sickness reducing manpower and excluding some lands from agricultural activities.

Threat to genetic resources

For a long time, it was considered that indigenous West African *Bos taurus* cattle were less productive and farmers, looking for higher producing animals, do not hesitate to cross these animals with *Bos indicus* which in general are larger animals but susceptible to trypanosomosis. Governments in the early sixties supported the crossbreeding with European dairy cattle breeds aimed at the increase of milk production. These practices expose indigenous cattle breeds, and particularly the West African *Bos taurus*, to the risk of disappearing following the uncontrolled upgrading by *Bos indicus* and other higher producing breeds which are susceptible to trypanosomosis. In the early eighties the Food and Agriculture Organization of the United Nations started alerting public opinion on animal genetic resources conservation and management (FAO 1981) to remedy this situation. At the earth summit in 1992 in Rio de Janeiro, 157 countries signed the United Nations convention on biological diversity and subsequently FAO (1992) launched a special program for the global management of farm animal genetic resources. The aim of this program is to maintain a maximum genetic diversity in context of gene pool for each species to allow for future unforeseen needs in the development of sustainable animal production systems. In addition the program aims at prompt actions to preserve animal breeds at risk of extinction.

ILCA/FAO/UNEP (1979), based on a survey in 18 western and central African countries, found that productivity of N'Dama and West African Shorthorn cattle was only marginally below that of zebu in terms of body weight, calving rate and mortality in low tsetse-infested areas. Recent studies have shown that under low trypanosomes challenge, zebu has higher absolute production than West African humpless cattle, but when this production is reported per unit of input or relative to the size of the animal, humpless animals have a better production index (Agyemang et al. 1991; Uza 1997). Mean growth rates of N'Dama and Kenyan zebu Boran are similar under good management and absence of trypanosome infection. In contrast, in presence of trypanosome infection, N'Dama does not require any treatment and in addition has higher weight gain than Boran. The belief that N'Dama and West African Shorthorn cattle are of small size and therefore are less productive is now obsolete.

Evidence of genetic basis for trypanotolerance

N'Dama and West African Shorthorn cattle (Baoulé, Muturu, Lagune) are recognized for their ability to withstand the effect of trypanosomes infection and to remain productive in areas where trypanosomosis prevents the presence of other cattle types, or significantly

reduces their productivity (Murray and Trail 1984; Trail et al. 1989). The ability to withstand trypanosome infection, known as trypanotolerance, has been under-exploited as means of trypanosomosis control. Trypanotolerance seems to be the result of a long adaptation through natural selection of *Bos taurus* introduced to the continent five to seven thousand years ago (ILRAD 1989). Many experimental studies have shown the trypanotolerant character of N'Dama and West African Shorthorn cattle compared to zebu and European breeds (e.g. Murray et al. 1982; Logan et al. 1988; Lorenzini et al. 1988; Pinder et al. 1988; Claxton and Leperre 1991; Doko et al. 1991; Dwinger et al. 1992). Some questions, however, persist about whether or not a previous infection with the parasites could have an effect on the results of the experiments. Further it is difficult to assure that different groups of animals undergo uniform infective challenge when they are exposed to field challenge or wild caught tsetse. An experiment carried out at the International Livestock Research Institute (ILRI) in Nairobi-Kenya (Paling et al. 1991), cleared doubts about these questions.

The experiment was based on embryos collected from N'Dama donors in The Gambia and implanted into recipient Boran mothers in Kenya (Jordt et al. 1986). The N'Dama cattle generated from this embryo transfer and challenged under controlled conditions (virulence of the trypanosome clone, infectivity of the tsetse, high ad-libitum plane of nutrition) confirmed their superior tolerance compared to Boran cattle. The superiority was in terms of controlling the parasitaemia, the ability to resist to anemia and the ability to produce more effective immune response to the trypanosome. The question is which of these characteristics can be used as indicator for trypanotolerance. More recent studies proved that the ability to resist anemia measured by packed cell volume (PCV) and the ability to control parasitaemia can be reliable criteria that can be used to evaluate trypanotolerance in cattle (Paling et al. 1991; Murray et al. 1990; Trail et al. 1990). The abilities to resist anemia and to control parasitaemia, although under genetic control, are not directly correlated to each other but both are correlated to productivity (Paling et al. 1991). The mechanisms by which trypanotolerant cattle control anemia and parasitaemia are not understood yet, but humoral immune response to trypanosome infection seems to be the major feature to define trypanotolerance in cattle (d'Ieteren et al. 1998).

It has now been definitely established that the trypanotolerance is an innate and heritable trait within N'Dama population, but its stability is influenced by environmental factors. The most important environmental factor appears to be the nutritional status after a study reported by Agyemang et al. (1990). This study showed that when supplemented with small amounts of groundnut meal, a local by-product in a village in The Gambia, infected N'Dama cattle not only have their growth rates improved but they were significantly less anemic than non-supplemented animals.

The genetic characteristic of trypanotolerance and the production efficiency observed in N'Dama cattle suggest that an improvement can be achieved for both traits by means of traditional selection program. A selection index including tolerance traits and production traits may be a way to value the breed. Before that, studies should focus on the determination of genetic parameters. Such selection program, aiming at improving the indigenous cattle breeds, will be much more sustainable than a crossbreeding program with exotic breeds.

Current research

Progress in molecular biology provides new tools for research in trypanosomosis and trypanotolerant animals. The bottleneck due to the ability of the parasites to undergo antigenic variations and thus escape the host immune responses during infection is expected to be overcome. With these new tools, some investigations are underway aiming at better diagnosing trypanosomosis and at developing vaccines against both parasites and disease, and maybe against the vector.

Diagnosis

Deoxyribonucleic acid (DNA) probes and polymerase chain reaction (PCR) technique are now used to diagnose trypanosome infection (Majiwa 1989; Nyeko et al. 1990; Majiwa and Sones 1993; Duvallet et al. 1993; Eshita et al. 1998). Reifenberg et al. (1997), using PCR to identify trypanosomes both in tsetse fly and in cattle in southern Burkina-Faso, found that the method has high potential to detect cryptic and mixed infections of trypanosomes in different hosts. They also reported that the method differentiated between the riverine forest type and the savanna type of *Trypanosoma congolense*. From blood spotted on filter papers and using PCR technique, Katakura et al. (1997) succeeded to detect 12% more animals infected with *T. congolense* and 24% more animals infected with *T. brucei* than the method of thin blood smear realized on animals positive for the microhaematocrit method. Clausen et al. (1998) compared the sensitivity and specificity of PCR with the haematocrit centrifugation technique and the mini-anion-exchange centrifugation technique and found that the detection rate of PCR is about twice as high as the direct parasitological techniques. In the future it may be possible to detect and use specific proteins for the diagnosis of trypanosomes and their infections.

Vaccine against parasites

The use of molecular biology tools in the diagnosis of the parasite as well as in the knowledge of the structure and the physiology of trypanosomes is underway. Studies on different parasite constituents (proteins, mRNA, genomic DNA) may provide in the near future some candidate antigens from which successful vaccines can be developed (Masake et al. 1997; Urakawa et al. 1997; Djikeng et al. 1998). Vaccine development based on variant surface glycoprotein has become less attractive because of the antigenic variation of trypanosomes (d'Ieteren et al. 1998). More attention is now being given to invariant surface proteins that are restricted to the flagellar pocket of the trypanosome (e.g., Nolan et al. 1999).

Vaccine against diseases

The idea behind this vaccine development is to hamper the mechanism by which trypanosomes cause the disease. It is reported that trypanosomes release factors that modulate the immune system of the host and disturb the normal function of the organism (Darji et al. 1996; Vaidya et al. 1997). The neutralization of these factors should allow the normal physiological function of the host defense system. An example of such factors is the cysteine protease, namely congopain, isolated from bloodstream forms of *T. congolense* (Mbawa et al. 1992; Authié et al. 1992) and found also in the blood of infected cattle (Authié et al. 1993a).

Trypanotolerance in cattle

Trypanotolerance is an innate feature of the long horn N'Dama and other Shorthorn cattle from West Africa (Roberts and Gray 1973; Roelants 1986; Doko et al. 1991). A problem related to trypanotolerance remains the trait definition and its measurement. Elucidating the mechanisms underlying the ability of trypanotolerant cattle to control the development of parasites and to limit anemia would give means to characterize trypanotolerance (d'Ieteren et al. 1998). Differences in immunological responses, however, have been observed between N'Dama and Boran following parasites challenge. A group of IgM of low specificity has been found in susceptible Boran whereas it was absent in N'Dama. Conversely high and sustained levels of trypanosome-specific IgG1 isotype were detected in N'Dama whereas the levels were low and transient in Boran (Authié et al. 1993b). In addition it has been found that the sialic acid content of erythrocytes decreases during trypanosome infection causing abnormal aging of the cells which were subsequently cleared by phagocytic cells contributing to anemia (Esievo et al. 1982). Trypanotolerant cattle were found to have more sialic acid in their red cells (Esievo et al. 1986) and were

able to maintain higher complement levels during trypanosome infection (Authié and Pobel 1990) than zebu cattle. The role of the different immunological responses in the control of trypanosome infections and the disease is not known and is being addressed by current research.

Under moderate challenge, Trail et al. (1991b) reported on some genetic parameters such as heritability for growth, average PCV and lowest PCV, and the genotypic and phenotypic correlations between these traits during a three-month testing period on 148 one-year-old N'Dama. The heritability of growth was 0.39, that of average PCV was 0.64 and that of lowest PCV reached during the test was 0.50. The genetic correlation between growth and average PCV was estimated at 0.70 and between growth and lowest PCV was 0.28. These estimates had high standard errors resulting from the small number of sires and offspring available per sire. These results indicate, however, that there are possibilities for improving the breed performance through a selection program.

Trypanotolerance at first sight can be seen as a binomial trait meaning that an animal is tolerant or not tolerant. A close look at the tolerance character shows that there is great variation within the recognized N'Dama breed which fails to withstand the disease under high tsetse-infected pressure (Ansell 1985). Van der Waaij et al. (2000) developed a threshold model for resistance (tolerance) where the animals are not productive below a minimal value of resistance and fully productive above a maximal value of resistance. Between the two values of resistance the production is defined as a function of resistance. This approach is used to assess consequences of various selection schemes.

Trypanotolerance-related genes and quantitative trait loci (QTL) detection

At ILRI in Nairobi (Kenya), a linkage analysis is underway on a F_2 generation of a cross between N'Dama and Boran cattle to identify genes or QTL involved in trypanotolerance (Teale 1993). The detection and identification of genes (or QTL) related to trypanotolerance offers new perspectives. These perspectives include selection program using genetic markers known as marker-assisted selection (Kashi et al. 1990), creation of transgenic animals, and QTL introgression using genetic markers known as marker-assisted introgression (MAI). Policy to introgress cattle with trypanotolerance genes should pay attention to future environment impacts of this introgression.

Marker-assisted introgression

Introgression is a crossbreeding system that introduces a beneficial allele from a donor line into a recipient line (Soller and Plotkin-Hazan 1977; Groen and Smith 1995). It generally

consists of three phases: first the production of F1 progeny by crossing donor and recipient lines. Second, several generations of backcrossing of the crossbred progeny to the recipient line while maintaining the beneficial allele from the donor line. Finally a third phase in which individuals from the last backcross are intercrossed to establish a new line that is homozygous for the desired allele from the donor line. The scope and interest for introgression are increasing as more genes of major importance become known (Smith et al. 1987; Kinghorn et al. 1994). The backcrossing phase is used to recover as much as possible the alleles at the other loci from the recipient line. During this backcrossing phase, it is very important that the beneficial allele can be recognized before reproductive age. In some cases, for example coat color, the allele can be recognized by phenotype. In a large number of cases, however, the recognition of the allele by phenotype in not possible. In those cases genetic markers that are linked to the gene of interest can be used. Genotypes for genetic markers can be determined in both sexes and at a young age. There is also possibility to identify those genetic markers at an earlier stage of the progeny's life i.e. on the fetus (Georges and Massey 1991).

Experiments in cattle to evaluate the efficiency of MAI are not feasible at this stage due not only to the lack of markers linked to trypanotolerance, but also to the time and money needed to perform such experiments. The existence of variation in resistance to *T. congolense* in some strains of laboratory mice (Morrison and Murray 1979) offers opportunity to undertake such an experiment which is the core subject of present thesis.

Mouse model

Among the laboratory mouse strains showing variation in resistance to *T. congolense*, the C57BL/6 strain appears the most resistant whereas the A/J strain shows the lowest level of resistance (Morrison et al. 1978). Although the survival times vary considerably between strains following challenge, the infection with *T. congolense* usually results in death of all strains of mice denoting the relative character of resistance to trypanosomosis (Teale et al. 1999). Based on two F_2 populations of mice, Kemp et al. (1996 &1997) show that three chromosomal regions are associated with trypanotolerance in mice. Following challenge with *T. congolense*, Iraqi et al. (2000) succeed to map QTL on chromosome 17 influencing survival time in mice to a small confidence interval using an F_6 advanced intercross line as described by Darvasi and Soller (1995).

The dense markers now available on the mouse genome and the identification of markers linked to these QTL enable experiments to evaluate consequences of marker-assisted introgression in mice. In 1997, it was decided to initiate such an experiment. This introgression experiment aims at the assessment of the effectiveness of MAI. This

experiment is expected to provide important information for the implementation of marker-assisted introgression in livestock.

Aim and outline of this thesis

The overall aim of this thesis is to develop optimum breeding schemes for the introgression of trypanotolerance genes using the mouse as a model organism.

Prior to QTL introgression, detection and location of genetic markers linked to that QTL are required. Methods to detect and locate QTL have been proposed exploiting crosses between inbred lines. The exploited underlying characteristics are recombination events that occur during meioses through generations. In chapter 2, proportions of recombinant haplotypes are assessed in an experimental F_9 population of mice, based on genotyping data for the ninth generation. These proportions are estimated using a direct counting method, a maximum likelihood method and CRIMAP (Lander and Green 1987). Furthermore, a simulation study is performed to assess the effect of misclassification in genotyping on estimated recombination fractions using maximum likelihood method.

Chapter 3 proposes methods to predict the amount of genotyping required during a backcross experiment in order to produce a specified number of individuals with the desired genotypes. Major factors identified and studied are the number of founder animals, the number of times each male is used (corresponding to the mating ratio), the number of backcross generations, the number and size of the regions to be introgressed and the influence of risk limitation.

The optimization of introgression strategies targeting three unlinked QTL (Chapter 4) aims at decreasing the total number of animals to be genotyped given a specified number of individuals homozygous at three QTL at the end of the introgression program. First, the optimization considers backcross and intercross phases separately and develops some alternatives for each phase. Second, it studies combinations of backcross and intercross alternatives as introgression strategies. Introgression strategies are then compared with respect to the number of animals to be genotyped, the number of genotypings and the costs.

In chapter 5, a marker-assisted introgression experiment involving two inbred lines (A/J susceptible and C57BL/6 resistant) has been set up to study the effectiveness of an introgression program. The experiment started with a reciprocal crossing of the two parental lines producing an F_1 followed by four generations of backcrossing to the susceptible strain, an intercrossing phase and a multiplication phase. During the backcrossing and intercrossing phases, animals were selected based on markers flanking the identified QTL. Ten groups of animals carrying different combinations of marker-

genotypes were produced. Theses groups and the two parental lines were challenged with *Trypanosoma congolense*. Trypanotolerance was measured as the survival time following the challenge over a period of 13 weeks.

Chapter 6 presents an overview of the contribution of the previous chapters to introgression using genetic markers. Further it describes opportunities for the introgression of trypanotolerance genes in cattle according to current reproduction technologies (natural service, artificial insemination (AI), AI combined with MOET). Chapter 6 also reviews production systems in Africa with an emphasis on livestock as well as breeding systems. Finally, constraints and future opportunities of such an introgression of trypanotolerance genes in cattle in sub-Saharan Africa are discussed in the last section.

References

Agyemang K, Dwinger RH, Touray BN, Jeannin P, Fofana D, Grieve AS (1990) Effect of nutrition on degree of anaemia and liveweight changes in N'Dama cattle infected with trypanosomes. *Livest Prod Sci* 26, 39-51

Agyemang K, Dwinger RH, Grieve AS, Bah ML (1991) Milk production characteristics and productivity of N'Dama cattle kept under village management in the Gambia. *J Dairy Sci* 74, 1599-1608

Ansell R (1985) Cattle breeding in the tropics. *World Anim Rev* 54, 30-38

Authié E, Pobel T (1990) Serum haemolytic and C3 levels in bovine trypanosomiasis under natural conditions of challenge – early indication of individual susceptibility to disease. *Vet Parasitol* 35, 43-59

Authié E, Muteti DK, Mbawa Z, Lonsdale-Eccles J, Webster P, Wells C (1992) Identification of a major antigen of *Trypanosoma congolense* as a cysteine protease. *Molec Biochem Parasit* 56, 103-116

Authié E, Duvallet G, Robertson J, Williams DJL (1993a) Antibody responses to a cysteine protease of *Trypanosoma congolense* – relationship to trypanotolerance in cattle. *Parasite Immunol* 15, 465-474

Authié E, Muteti DK, Williams DJL (1993b) Antibody responses to non-variant antigens of *Trypanosoma congolense* in cattle of differing susceptibility to trypanosomiasis. *Parasite Immunol* 15, 101-111

Brown CGD, Hunter AG, Luckins AG (1990) Diseases caused by protozoa. In: MMH Sewell and DW Brocklesby (Eds), *Handbook on animal diseases in the tropics,* (4th Ed). Bailliere Tindall, London, pp.161-226

Clausen PH, Wiemann A, Patzelt R, Kakaire D, Poetzsch C, Peregrine A, Mehlitz D, Jongejan F, Goff W, Camus E, (1998) Use of a PCR assay for the specific and sensitive detection of *Trypanosoma spp.* in naturally infected dairy cattle in peri-urban Kampala, Uganda. *In Tropical Veterinay Medicine* (Jongejan F and Goff W, eds): *molecular epidemiology, hemoparasites and their vectors, and general topics.* Annals of the New York Academy of Sciences 849, 21-31

Claxton JR, Leperre P (1991) Parasite burdens and host susceptibility of zebu and N'Dama cattle in village herds in The Gambia. *Vet Parasitol* 40, 293-304

Darvasi A, Soller M, (1995) Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* 141, 1199-1207

Darji A, Beschin A, Sileghem M, Heremans H, Brys L, de Baetselier P (1996) In vitro simulation of immunosuppression caused by *Trypanosoma brucei*: active involvement of gamma interferon and tumor necrosis factor in the pathway of suppression. *Infect Immun* 64, 1937-1943

d'Ieteren GDM, Authié E, Wissocq N, Murray M (1998) Trypanotolerance, an option for sustainable livestock production in areas at risk from trypanosomosis. *Rev Sci Tech OIE* 17, 154-175

Djikeng A, Agufa C, Donelson JE, Majiwa PA (1998) Generation of expressed sequence tags as physical landmarks in the genome of *Trypanosoma brucei*. *Gene* 221, 93-106

Doko A, Guedegbe B, Baelmans R., Demey F, N'Diaye A, Pandey VS, Verhulst A (1991) Trypanosomiasis in different breeds of cattle from Benin. *Vet Parasitol* 40, 1-7

Duvallet G, Bengaly Z, Reifenberg JM, Argiro L, Sones KR (1993) New tools for diagnosis and epidemiology of African animal trypanosomiasis: preliminary results at CRTA/CIRDES. *22nd meeting of International Scientific Council for Trypanosomiasis Research and Control (ISCTRC) held in Kampala, Uganda.* ISCTRC 117, 59-67

Dwinger RH, Clifford D, Agyemang K, Gettingy G, Grieve AS, Kora S, Bojang MA (1992) Comparative studies on N'Dama and zebu cattle following repeated infections with *Trypanosoma congolense*. *Res Vet Sci* 52, 292-298

Esievo KAN, Saror DI, Ilemobade AA, Hallaway MH (1982) Variation in erythrocyte surface and free serum sialic acid concentrations during experimental *Trypanosoma vivax* infection. *Res Vet Sci* 32, 1-5

Esievo KAN, Saror DI, Kolo MN, Eduvie LO (1986) Erythrocyte surface sialic acid concentrations in N'Dama and Zebu cattle. *J Comp Pathol* 96, 95-99

Eshita Y, Majiwa PA, Urakawa T, Inoue N, Hirumi K, Yanagi T, Yoneda Y, Hara T, Higuchi T, Fukuma, T, Hirumi H (1998) The application of molecular biological tools to epidemiology of African trypanosomiasis. *Tokai J Exp Clin Med* 23, 401-411

FAO (1981) Animal genetic resources conservation and management. Proceedings of the FAO/UNEP technical consultation. FAO Animal Production and Health paper 24. FAO, Rome, 388pp

FAO (1992) The management of global animal genetic resources. FAO Animal Production and Health paper 104. FAO, Rome, 309p

Georges M, Massey JM (1991) Velogenetics, or the synergistic use of marker assisted selection and germ-line manipulation. *Theriogenology* 35, 151-159

Groen AF, Smith C (1995) A stochastic simulation study of the efficiency of marker-assisted introgression in livestock. *J Anim Breed Genet* 112, 161-170

ILCA/FAO/UNEP (1979) Trypanotolerant livestock in West and Central Africa. ILCA monograph no. 2. ILCA, Addis-Ababa, Ethiopia, 303 pp

ILRAD (1989) N'Dama cattle: managing Africa's genetic resource. Reports/ILRAD, October 1989

Iraqi F, Clapcott SJ, Kumari P, Haley CS, Kemp SJ, Teale AJ (2000) Fine mapping of trypanosomiasis resistance loci in murine advanced intercross lines. *Mamm Genome* 11, 645-648

Jordt T, Mahon GD, Touray GN, Ngulo WK, Morrison WI, Rawle J, Murray M (1986) Successful transfer of frozen N'Dama embryos from The Gambia to Kenya. *Trop Anim Health Pro* 18, 65-75

Kashi Y, Hallerman E, Soller M (1990) Marker assisted selection of candidate bulls for progeny testing programmes. *Anim Prod* 51, 63-74

Katakura K, Lubinga C, Chitambo H, Tada Y (1997) Detection of *Trypanosoma congolense* and *T. brucei* subspecies in cattle in Zambia by polymerase chain reaction from blood collected on a filter paper. *Parasitol Res* 83, 241-245

Kemp SJ, Darvasi A, Soller M, Teale AJ (1996) Genetic control of resistance to trypanosomiasis. *Vet Immunol Immunop* 54, 239-243

Kemp SJ, Iraqi F, Darvasi A, Soller M, Teale AJ (1997) Localization of genes controlling resistance to trypanosomiasis in mice. *Nat Genet* 16, 194-196

Kinghorn B, van Arendonk JAM, Hetzel J (1994) Detection and use of major genes in animal breeding. *AgBiotech News and Information* 6, 297-302

Kristjanson PM, Swallow BM, Rowlands GJ, Kruska RL, de Leeuw PN (1999) Measuring the costs of African animal trypanosomiasis, the potential benefits of control and returns to research. *Agr Syst* 59, 79-98

Lander EP, Green P (1987) Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci* 84, 2363-2367

Lessard P, L'Eplattenier R, Norval RAI, Dolan TT, Croze H, Kundert K, Walker JB, Irvin AD, Perry BD (1990) Geographical information systems for studying the epidemiology of cattle diseases caused by *Theileria parva*. *Vet Rec* 126, 255-262

Logan LL, Paling RW, Moloo SK, Scott JR (1988) Comparative studies on the responses of N'Dama and Boran cattle to experimental challenge with tsetse-transmitted *Trypanosoma congolense*. In: *Livestock production in tsetse affected areas of Africa. Proceedings of a meeting held 23-27 November 1987 in Nairobi, Kenya*. ILCA/ILRAD, Nairobi, Kenya, pp. 152-160

Lorenzini E, Scott RJ, Paling RW, Jordt T (1988) The effect of *Trypanosoma congolense* infection on the reproductive cycle of N'Dama and Boran heifers. In: *Livestock production in tsetse affected areas of Africa. Proceedings of a meeting held 23-27 November 1987 in Nairobi, Kenya*. ILCA/ILRAD, Nairobi, Kenya, pp. 68-73

MacLennan KJR (1980) Tsetse-transmitted trypanosomiasis in relation to the rural economy in Africa: Part 1. Tsetse infestation. *World Anim Rev* 36, 2-17

Majiwa PAO (1989) Recombinant DNA probes as tools for epidemiological studies of parasitic diseases. *Discovery and Innovation* 1, 35-40

Majiwa PAO, Sones DR (1993) DNA probe and PCR-based methods for the detection of trypanosomes. *22nd meeting of International Scientific Council for Trypanosomiasis Research and Control (ISCTRC) held in Kampala, Uganda*. ISCTRC 117, 53-58

Masake RA, Majiwa PA, Moloo SK, Makau JM, Njuguna JT, Maina M, Kabata J, ole-MoiYoi OK, Nantulya VM (1997) Sensitive and specific detection of *Trypanosoma vivax* using the polymerase chain reaction. *Exp Parasitol* 85, 193-205

Mbawa ZR, Gumm ID, Shaw E, Lonsdale-Eccles JD, (1992) Characterisation of a cysteine proteinase from bloodstream forms of *Trypanosoma congolense*. *Eur J Biochem* 204, 371-379

Morrison WI, Roelants GE, Mayor-Withey KS, Murray WI (1978) Susceptibility of inbred strains of mice to *Trypanosoma congolense*: correlation with changes in spleen lymphocyte populations. *Clin Exp Immunol* 32, 25-40

Murray M, Morrison WI, Whitelaw DD (1982) Host susceptibility to African trypanosomiasis: trypanotolerance. *In Advances in parasitology, Vol. 21 (J.R. Baker and R. Muller, eds).* Academic Press, London, 1-68

Murray M, Trail JCM (1984) Genetic resistance to animal trypanosomiasis in Africa. *Prev Vet Med* 2, 541-551

Murray M, Trail JCM, Davis CE, Black SJ (1984) Genetic resistance to African trypanosomiasis. *J Infect Dis* 149, 311-319

Murray M, Trail JCM, d'Ieteren GDM (1990) Trypanotolerance in cattle and prospects for the control of typanosomiasis by selective breeding. *Rev Sci Tech OIE* 9, 369-386

Nolan DP, Geuskens M, Pays E (1999) N-linked glycans containing linear poly-N-acetyllactosamine as sorting signals in endocytosis in *Trypanosoma brucei. Current Biology* 9, 1169-1172

Nyeko JHP, Ole-Moiyoi OK, Majiwa PAO, Otieno LH, Ociba PM (1990) Characterization of trypanosome isolates from cattle in Uganda using species-specific DNA probes reveals predominance of mixed infections. *Insect Science and its Application* 11, 271-280

Paling RW, Moloo SK, Scott JR, Gettinby G, McOdimba FA, Murray M (1991) Susceptibility of N'Dama and Boran cattle to sequential challenges with tsetse-transmitted *Trypanosoma congolense. Parasite Immunol* 13, 427-445

Paling RW, Dwinger RH (1993) Potential of trypanotolerance as a contribution to sustainable livestock production in tsetse affected Africa. *Vet Quart* 15, 60-67

Pinder M, Bauer J, van Melick A, Fumoux F (1988) Immune responses of trypanoresistance and trypanosusceptible cattle after cyclic infection with Trypanosoma congolense. *Vet Immunol Immunop* 18, 245-275

Reifenberg JM, Solano P, Bauer B, Kabore I, Cuny G, Duvallet G, Cuisance D (1997) Advantage of the PCR technique in assessing the epidemiology of bovine trypanosomiasis: example of the agropastoral development area of Yale, Burkina-Faso. *Rev Elev Med Vet Pays Trop* 50, 14-22

Roberts CJ, Gray AJ (1973) Studies on trypanosome-resistant cattle II. The effects of trypanosomiasis on N'Dama, Muturu and Zebu cattle. *Trop Anim Health Pro* 5, 220-233

Roelants GE (1986) Natural resistance to African trypanosomiasis. *Parasite Immunol* 8, 1-10

Savadogo M, (2000) Crop residue management in relation to sustainable land use: a case study in Burkina Faso. PhD dissertation, Wageningen University, The Netherlands

Smith C, Meuwissen THE, Gibson JP (1987) On the use of transgenes in livestock improvement. *Anim Breed Abstracts* 55, 1-10

Soller M, Plotkin-Hazan J (1977) The use marker alleles for the introgression of linked quantitative alleles. *Theor Appl Genet* 51, 133-137

Teale AJ (1993) Improving control of livestock disease. Animal biotechnology in the Consultative Group for International Agricultural Research. *Bioscience* 43, 475-483

Teale A, Agaba M, Clapcott S, Gelhaus A, Haley C, Hanotte O, Horstmann R, Iraqi F, Kemp S, Nilsson P, Schwerin M, Sekikawa K, Soller M, Sugimoto Y, Womack J (1999) Resistance to trypanosomiasis: of markers, genes and mechanisms. *Arch Tierz Dummerstorf* 42, 36-41

Trail JCM, d'Ieteren GDM, Teale AJ, (1989) Trypanotolerance and the value of conserving livestock genetic resources. *Genome* 31, 805-812

Trail JMC, d'Ieteren GDM, Feron A, Kabiese O, Mulungo M, Pelo M (1991a) Effect of trypanosome infection, control of parasitaemia and control of anaemia development on productivity of N'Dama cattle. *Acta Trop* 48, 37-45

Trail JMC, d'Ieteren GDM, Murray M (1991b) Practical aspects of developing genetic resistance to trypanosomiasis. In Axford, R.F.E. and Owen, J.B. (eds): *Breeding for disease resistance in farm animals*. CAB International, Wallington, pp. 224-234

Urakawa T, Eshita Y, Majiwa PA (1997) The primary structure of *Trypanosoma (Nannomonas) congolense* variant surface glycoproteins. *Exp Parasitol* 85, 215-224

Uza DV (1997) The productivity of Muturu cattle (*Bos brachyceros*) under ranching conditions in the Southern Guinea Savanna of Benue state, Nigeria. *Outlook on Agriculuture* 26, 19-23

Vaidya T, Bakhiet M, Hill KL, Olsson T, Kristensson K, Donelson JE (1997) The gene for a T lymphocyte triggering factor from African trypanosomes. *J Exp Med* 186, 433-438

Van der Waaij EH, Bijma P, Bishop SC, Van Arendonk JAM (2000) Modelling selection for production traits under constant infection pressure. *J Anim Sci* (*in press*)

CHAPTER II

Estimating recombination in an advanced intercross population of mice

Olorounto D. KOUDANDÉ[1], Fuad IRAQI[2], Johan A. M. Van ARENDONK[1] and Henk BOVENHUIS[1]

[1] *Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands*

[2] *International Livestock Research Institute (ILRI), P.O. Box 30709 Nairobi, Kenya*

Submitted for publication

Abstract

The present study aims to assess the proportion of recombinant haplotypes in an advanced intercross F_9 population of mice (A/J x C57BL/6), based solely on genotyping data collected from that population. The targeted chromosomal region of approximately 7-cM is located on chromosome 17 and contains a major QTL associated with trypanoresistance in mice. The estimation of the proportion of recombinant haplotypes is performed through a counting method and a maximum likelihood estimation (MLE) method that does and does not account for misclassifications. A simulation study is performed to evaluate the characteristics of these two methods. Finally CRIMAP program has been used on the same data set. Allele frequencies within the F_9 population show deviations from the expected frequency (0.5); these deviations can be explained by random drift. The counting method gives estimates for the proportion of recombinant haplotypes in F_9 that are biased upward. Results from the maximum likelihood model suggest 63% higher proportion of recombinant haplotypes whereas CRIMAP analysis doubles the value obtained from the mouse genome database. The current data set suggests a different marker order other than that reported by the mouse genome database. Misclassification is an important factor to be considered when analyzing such data even though the accuracy of the estimates of misclassified fraction is low. The results of this study clearly show evidence of a strong accumulation of recombinant haplotypes within the target chromosomal region in the F_9.

Introduction

Most of the studies to detect and map quantitative trait loci (QTL) in animals, and to a lesser extent in plants, have been performed using a F_2, backcross, half-sib or full-sib experimental design. The confidence intervals of QTL being mapped are generally large due to the relative small number of meiosis limiting the number of recombinant events in small chromosomal regions (Darvasi and Soller 1995). A number of experimental designs have been proposed to increase recombination events in crosses, like Recombinant Inbred Strains (RIS) (Bailey 1971), Recombinant Congenic Strains (RCS) (Demant and Hart 1986), Advanced Intercross Lines (AIL) (Darvasi and Soller 1995), and Interval Specific Congenic Strains (ISCS) (Darvasi 1997 & 1998). These designs are based either on a number of backcross generations followed by generations of brother-sister mating (for RCS), or on additional generations of semi-random intercrossing after the F_2 generation (for AIL). All these methods aim to increase recombination events on the genome allowing mapping QTL with a higher resolution. Nadeau et al. (2000) propose the design of panels of Chromosomes Substitution Strains (CSS) for QTL mapping. The CSS consists of a host

strain in which one entire chromosome is substituted by the donor's corresponding chromosome.

An AIL is created by producing F_1 animals from crossing two inbred lines followed by generations of random intercrossing. One advantage of this method for QTL mapping is that no genotyping and no phenotyping are needed during the process of intercrossing called 'breeding generations'. Darvasi and Soller (1995) derived a formula to predict the expected proportion of recombinant haplotypes as a function of the number of generations and the recombination fraction. Although the principle of accumulation of recombinant events is exploited for fine mapping of QTL, few reports are available on the evaluation of recombination in an AIL population. Talbot et al. (1999) succeeded to map a QTL influencing emotion in mice in a 0.8-cM interval from a heterogeneous stock. The heterogeneous stock was at generation 58 at the time of their experiment and was obtained 30 years earlier from an eight-way cross of inbred mouse strains. Iraqi et al. (2000) used two AIL populations (both in generation 6), originally produced by crossing two inbred lines of mice, to fine map a QTL for trypanoresistance in a 0.9-cM interval on chromosome 17.

The use of genotyping data can be subjected to errors or misclassifications. When analyzing milk protein genotypes, Bovenhuis and Van Arendonk (1991) estimated that the fraction of misclassified animals was 8-10%. These misclassifications included pedigree errors and genotyping errors. Estimates of fractions of misclassified animals had high standard errors. Misclassifications may be due to a faint amplification of the DNA, errors in reporting the genotyping results (for example, mixing up results from different animals) and can be corrected by repeated genotyping. Misclassifications resulting from errors in the pedigree, however, cannot be checked by repeated genotyping (Bovenhuis and Van Arendonk 1991), but they can be (partly) identified if pedigree information is available.

Darvasi and Soller (1995) stipulated that "for QTL mapping purposes, only individuals from one of the later generations are phenotyped and genotyped." In such a situation where pedigree information is not available, the present paper aims to assess the proportion of recombinant haplotypes in an experimental F_9 population of mice. The targeted chromosomal region, of approximately 7-cM according to mouse genome database (MGD 1999), is located on chromosome 17 and recognized as containing a major QTL associated with trypanoresistance in mice (Kemp et al. 1997). The F_9 population is the follow up of the F_6 (A/J x C57BL/6) used by Iraqi et al. (2000). The proportions of recombinant haplotypes are first estimated using a direct counting method and a maximum likelihood estimation method (MLE). A simple MLE is primary developed followed by an MLE accounting for misclassification. A simulation study is performed to assess the characteristics of these two methods. Further the proportion of recombinant haplotypes and

the markers order are estimated using the multi-point analysis of CRIMAP (Green et al. 1990).

Materials and methods

Mouse strains and generation of the advanced intercross population

Inbred mouse strains A/J (OlaHsdnd) and C57BL/6 (OlaHsd) purchased from Harlan UK Ltd., Bicester, U.K., are crossed reciprocally to produce the F_1 generation. At each of the following intercross generations, a group of new parents is formed by randomly selecting one male and one female from each litter. In the new group of parents, random mating is applied avoiding brother-sister mating. From F_1 through F_9, 50 males and 50 females are targeted as parents each generation. To achieve this goal, 55-60 pairs of animals are selected each generation to compensate eventual mortality and infertility problems. In that way, the expected increase of inbreeding is kept constant at a rate of 0.25 percent each generation (Falconer and Mackay 1996). At the ninth generation, several litters are produced per pair of selected F_8 animals to increase the number of progeny. A group of 400 F_9 mice is genotyped for the present study.

DNA extraction, microsatellite markers and genotyping

DNA is extracted from blood samples by conventional method described by Sambrook et al. (1989). The obtained DNA is diluted and quantified using a spectrophotometer (GeneQuant II, LKP Pharmacia Biochem, England, 1994) and each DNA solution is adjusted to 0.05 mg/ml for polymerase chain reaction (PCR).

Twelve fluorescent labeled microsatellite markers (Research Genetic Inc., Huntsville, AL, USA) are used to assess the genotypes of each animal at corresponding loci on the target region of chromosome 17 (see Table 2.2). The average marker spacing in this section of the chromosome is 0.64 cM. All 12 microsatellites are fixed at alternative alleles for the two mouse strains used to generate the AIL population under study.

The amplification of these microsatellite markers is performed according to the supplier recommendations using a thermocycler PTC100 (MJ Research, Inc., USA). PCR products are analyzed on a 4.25% polyacrylamide gel (Laemmli 1970) using an automated DNA sequencer ABI 377 (Perkin Elmer). Subsequently the tracking of the gel is checked and adjusted manually and analyses are performed using Genescan™ version 2.1 and Genotyper™ version 2.0 software (Applied Biosystems, 1996). The production of advanced intercross population and the genotyping are executed at the International Livestock Research Institute (ILRI), Nairobi, Kenya.

Data analysis

Expectation

Darvasi and Soller (1995) predicted the expected proportion of recombinant haplotypes (θ_t) in generation t between two loci as

$$\theta_t = \frac{1-(1-\theta)^{t-2}(1-2\theta)}{2} \qquad [1]$$

where θ is the recombination fraction. Map distances are transformed in recombination fractions using Haldane's (1919) mapping function. Expected proportions of recombinant haplotypes in the F_9 (θ_9) are calculated by applying Eq [1]. Three methods are used to assess the recombination parameters from the genotyping data, that is, a counting method, a maximum likelihood approach and the use of CRIMAP.

Counting method

To estimate the proportion of recombinant haplotypes (θ_t), the counting method can be used. The alleles coming from the inbred line A/J are denoted A and the alleles from the inbred line C57BL/6 are denoted C. Consequently haplotypes of type A-A are non recombinant haplotypes originating from line A/J as well as haplotypes of type C-C from line C57BL/6. According to the genotypes at two loci, four classes of animals can be distinguished and the method consists of counting the number of recombinant haplotypes within a marker bracket as follows:

Class_1 consists of animals that have genotypes AA for locus 1 and AA for locus 2 or CC for locus 1 and CC for locus 2; the haplotype configurations are A-A and A-A or C-C and C-C. These haplotypes are non-recombinant haplotypes and therefore zero recombinant haplotypes are counted for this class of animals.

Class_2 consists of animals that are homozygous at one locus and heterozygous at the other: AA for locus 1 and AC for locus 2, or AC for locus 1 and AA for locus 2 or CC for locus 1 and AC for locus 2, or AC for locus 1 and CC for locus 2. In each of these cases one recombinant and one non-recombinant haplotype is present and therefore one recombinant haplotype is counted.

Class_3 consists of animals that have genotypes AA for locus 1 and CC for locus 2 resulting in haplotypes A-C and A-C or animals with genotypes CC at locus 1 and AA at locus 2 resulting in haplotypes C-A and C-A. In this case two recombinant haplotypes are counted for each animal in this class.

Class_4 consists of animals that are heterozygous for both loci; AC for locus 1 and AC for locus 2. In this case, haplotypes cannot be determined without pedigree information and therefore these animals are excluded from the counting method.

The proportion of recombinant haplotypes ($\hat{\theta}_t$) is estimated by dividing the total number of recombinant haplotypes over twice the number of animals contributing to the estimate;

$$\hat{\theta}_t = \frac{(0 * N_1) + (1 * N_2) + (2 * N_3)}{2 * (N_1 + N_2 + N_3)} \qquad [2]$$

where N_1 is the number of animals in class_1, N_2 the number of animals in class_2 and N_3 the number of animals in class_3.

Maximum likelihood estimation (MLE)

Table 2.1 shows the nine groups of animals that can be distinguished according to their genotypes in an AIL at generation 9. Assuming that allele frequencies for A and C are 0.5 and there is random union of gametes, probabilities for the genotype combinations for any two loci can be derived (Table 2.1). Using these probabilities, the likelihood function can be established as:

$$L(\theta_t) \propto \left[\tfrac{1}{4}(1-\theta_t)^2\right]^{(n_1+n_9)} * \left[\tfrac{1}{2}(1-\theta_t)\theta_t\right]^{(n_2+n_3+n_5+n_6)} * \left[\tfrac{1}{4}\theta_t^2\right]^{(n_7+n_8)} * \left[\tfrac{1}{2}(1-\theta_t)^2 + \tfrac{1}{2}\theta_t^2\right]^{(n_4)} \qquad [3]$$

where θ_t is the proportion of recombinant haplotypes in generation F_t as defined in Eq [1], and n_i the number of animals in group i (Table 2.1). The value of θ_9 ($t = 9$) that maximizes $\ln[L(\theta_9)]$ also maximizes $L(\theta_9)$ and is obtained by the downhill simplex approach (Nelder and Mead 1965). The optimization procedure is run a number of times using different starting values to make sure that the global maximum is reached. Using the same method, the recombination fraction (θ) is estimated from a data set of F_2 animals (44 animals) of the same cross A/J x C57BL/6 and this data set consists of genotyping results on two markers, *D17Mit45* and *D17Mit11*.

Maximum likelihood estimation accounting for misclassification (MLE+*M*)

In the present study, only F_9 individuals are genotyped. This does not allow verification of the genotypes making use of pedigree information and therefore misclassified genotypes cannot be excluded. To study this effect, the data is also analyzed using a maximum likelihood model that accounts for misclassification.

Table 2.1 Genotype probabilities of offspring according to haplotype frequencies

Haplotypes →		A – A	A – C	C – A	C – C
↓	Frequency	$\frac{1}{2}(1-\theta_t)$	$\frac{1}{2}\theta_t$	$\frac{1}{2}\theta_t$	$\frac{1}{2}(1-\theta_t)$
A – A	$\frac{1}{2}(1-\theta_t)$	$\frac{1}{4}(1-\theta_t)^2$ (n_1)	$\frac{1}{4}(1-\theta_t)\theta_t$ (n_2)	$\frac{1}{4}(1-\theta_t)\theta_t$ (n_3)	$\frac{1}{4}(1-\theta_t)^2$ (n_4)
A – C	$\frac{1}{2}\theta_t$	$\frac{1}{4}(1-\theta_t)\theta_t$ (n_2)	$\frac{1}{4}\theta_t^2$ (n_7)	$\frac{1}{4}\theta_t^2$ (n_4)	$\frac{1}{4}(1-\theta_t)\theta_t$ (n_5)
C – A	$\frac{1}{2}\theta_t$	$\frac{1}{4}(1-\theta_t)\theta_t$ (n_3)	$\frac{1}{4}\theta_t^2$ (n_4)	$\frac{1}{4}\theta_t^2$ (n_8)	$\frac{1}{4}(1-\theta_t)\theta_t$ (n_6)
C – C	$\frac{1}{2}(1-\theta_t)$	$\frac{1}{4}(1-\theta_t)^2$ (n_4)	$\frac{1}{4}(1-\theta_t)\theta_t$ (n_5)	$\frac{1}{4}(1-\theta_t)\theta_t$ (n_6)	$\frac{1}{4}(1-\theta_t)^2$ (n_9)

n_i = number of individuals within group i

There is not enough information in the data to estimate misclassification for both loci separately, therefore it is assumed that the probability (M) of misclassification is the same for both loci. After some algebraic derivations, the likelihood function accounting for misclassifications is as follows:

$$\begin{aligned} L(\theta_t) \propto & \left[\tfrac{1}{4}(1-M)^2(1-\theta_t)^2 + \tfrac{2}{16}\left(M^2+2M(1-M)\right)\right]^{(n_1+n_9)} \\ & * \left[\tfrac{1}{2}(1-M)^2(1-\theta_t)\theta_t + \tfrac{8}{16}\left(M^2+2M(1-M)\right)\right]^{(n_2+n_3+n_5+n_6)} \\ & * \left[\tfrac{1}{4}(1-M)^2\theta_t^2 + \tfrac{2}{16}\left(M^2+2M(1-M)\right)\right]^{(n_7+n_8)} \\ & * \left[\tfrac{1}{2}(1-M)^2\left((1-\theta_t)^2+\theta_t^2\right) + \tfrac{4}{16}\left(M^2+2M(1-M)\right)\right]^{(n_4)} \end{aligned} \qquad [4]$$

The model assumes a mixture of two distributions, one for correctly genotyped individuals and one for misclassified individuals. If an individual is assigned to the correct genotype the distribution is as described in Eq [3]. The distribution of misclassified individuals is according to population frequencies.

Replacing θ_t in Eq [3] and Eq [4] by Eq [1] gives directly an estimate of the recombination fraction θ.

Simulation study

Simulation is used to study the effect of the estimation procedure on the estimate of the recombination fraction (θ) in situations with and without misclassifications. The structure of the simulated breeding program is as follow: 50 inbred mice (25 males and 25 females) from one line are crossed with 50 inbred mice (25 males and 25 females) from another line fixated at alternative alleles to produce generation F_1. From this generation till F_8, one male and one female are randomly selected from each pair of parents to be parents for the next generation. The pairing of males and females is at random avoiding brother-sister mating. The number of intercross generations is 9, litter size is five and results are based on 10000 replicated simulations based on 350 animals in F_9. Three methods are used to analyze the simulated data, that is the counting method, the MLE and the MLE accounting for misclassification. Three different data sets are simulated; for the first data set no misclassification is simulated, for the second data set 10% misclassification is simulated for both loci, and for the third data set 5% misclassification is simulated for locus one and 15% for locus 2. The simulated recombination fraction (θ) is 0.10.

A chi-square test is used to determine deviations from Hardy-Weinberg equilibrium. The distribution of the test statistics is empirically determined using the simulation described above. Based on 400 animals produced at F_9 and on 50000 replicates, the critical point at 0.01 level is 18.23 and 10.45 at 0.05 level.

CRIMAP-based analysis

The use of CRIMAP requires genotypes from parents and offspring. In the current case where no information was available on parents, all parents were assumed to be heterozygous for all marker loci under study. The ideal situation for this analysis would be to have one big family with all 400 F_9 genotyped mice to optimize the estimation of θ_9. The software, however, does not support such a big family, therefore, four families of equal size were assumed. First, a two-point analysis is performed to compare the outcome with previous analysis, followed by a multi-point analysis to assess θ_9 and the order of markers.

Using CRIMAP, the recombination fraction (θ) is also estimated from the F_2 genotyping data obtained from the cross A/J x C57BL/6. Because of the small number of F_2 animals, only one family was considered for the CRIMAP analysis.

Results

The number of animals (n) for which marker genotype results are available varies among markers and is different from the original 400 genotyped animals (Table 2.2). This variation results from the fact that genotyping information is missing for some animals at some marker loci due to absence of PCR products. Table 2.2 shows the genotype frequencies in the F_9 from which allele frequencies are derived. The chi-square test shows that all marker-loci are in Hardy-Weinberg equilibrium except *D17Mit191* (P<0.01). The allele frequencies in the F_9 population show a significant increase in alleles inherited from A/J line except at *D17Mit198* and *D17Mit81*.

Table 2.2 Microsattelite markers used and their location, observed allele and genotype frequencies in F_9, A=A/J allele and C=C57BL/6 allele, n = number of animals with observations

Marker[a]	n	Marker[a] position (cM)	Alleles size[a] (bp)		Observed genotypes frequency			Observed allele frequency (A)[b]	χ^2 test for genotype frequencies
			A	C	AA	AC	CC		
D17Mit29	370	0.00	146	150	119	188	63	0.576	ns
D17Mit198	386	0.90	116	102	108	194	84	0.531	ns
D17Mit45	341	1.30	216	226	118	156	67	0.575	ns
D17Mit81	347	1.80	126	106	107	157	89	0.525	ns
D17Mit191	381	2.10	96	90	90	231	60	0.539	**
D17Mit175	395	2.60	122	110	139	196	60	0.600	ns
D17Mit16	388	3.00	92	118	141	181	66	0.597	ns
D17Mit214	362	3.60	126	108	132	173	57	0.604	ns
D17Nds3	384	3.96	95	103	141	191	52	0.616	ns
D17Mit148	385	5.29	96	94	128	199	58	0.591	ns
D17Mit233	373	5.80	110	122	120	195	58	0.583	ns
D17Mit11	338	6.85	160	176	116	147	75	0.561	ns

(**) P < 0.01; (ns) non significant at 5% level for a simulated chi-square test for Hardy-Weinberg equilibrium with one degree of freedom (Hedrick 1985).
a) Source: MGD (1999), Dietrich et al. (1996)
b) All allele frequencies deviate from the expected value of 0.5 (95% confidence interval), except *D17Mit198* and *D17Mit81*

Table 2.3 shows the results of the simulation study. If all individuals are correctly genotyped, the maximum likelihood estimate (MLE) of recombination fraction (0.103) is very close to the simulated value of 0.100. When simulating 10% misclassifications at both marker loci, the recombination fraction is overestimated (0.122). If the fraction of misclassified individuals differs between the two loci (5% and 15%) the estimated recombination fraction, averaged over the 10000 replicates, is also 0.122. The recombination fraction is underestimated with MLE accounting for misclassification. For

example when no misclassifications are simulated for both loci, the estimate of the recombination fraction is 0.087 and the estimated misclassified fraction is 0.074. With 10% misclassifications at both loci the estimated recombination fraction is 0.094. In this case the estimated fraction misclassified is 0.130. Estimates of misclassification show large empirical standard errors indicating that this parameter cannot be estimated accurately. The simulated data was also analyzed using the counting method. For the simulated data with no misclassifications, the estimate for θ_9 was 0.362 and for the data with 10% misclassifications the estimate was 0.390. A simulated recombination fraction of 10% is expected to result in a fraction of recombinant haplotypes of 0.309 in the F_9 (Eq [1]).

Table 2.3 Mean estimated recombination fraction and misclassified fraction based on 10000 replicated simulations of an F_9 advanced intercross population consisting of 350 animals. The simulated θ is 0.10 and empirical standard errors are in parentheses

Methods[a]	Simulated misclassification		Estimated[b]	
	Locus 1	Locus 2	*M*	θ
MLE	0.00	0.00	n.e.	0.103 (0.032)
	0.10	0.10	n.e.	0.122 (0.033)
	0.05	0.15	n.e.	0.122 (0.032)
MLE + *M*	0.00	0.00	0.074 (0.125)	0.087 (0.033)
	0.10	0.10	0.130 (0.158)	0.094 (0.041)
	0.05	0.15	0.130 (0.157)	0.094 (0.041)

a: MLE = maximum likelihood estimation, MLE + *M* = MLE accounting for misclassification

b: *M* = misclassified fraction, θ = recombination fraction, n.e. = not estimated

The proportion of recombinant haplotypes in F_9 (θ_9) within each of the eleven marker brackets have been evaluated and summarized in Table 2.4. The counting method gives estimates of the proportion of recombinant haplotypes that are much higher than the proportions that would be predicted based on Eq [1] and the map distances reported in the MGD (1999). The maximum likelihood method gives estimates that are lower than the estimates obtained using the counting method but still higher than the expected proportion of recombinant haplotypes. For the F_2 data, the maximum likelihood method gives a recombination fraction of 0.046 between markers *D17Mit45* and *D17Mit11* corresponding to a map distance of 4.8 cM. The corresponding map distance in MGD (1999) data is 5.5 cM. The estimated proportion of recombinant haplotypes (θ_9) between markers *D17Mit45* and *D17Mit11* in generation F_9 is 0.284. Based on a distance of 5.5 cM, the estimated θ_9 is, however, 47% higher than the expected proportion of recombinant haplotypes (0.193).

Accounting for misclassifications resulted in lower estimates for the proportion of recombinant haplotypes than estimates from simple MLE (Table 2.4). This result is in agreement with what would have been expected based on the simulation study. The estimated misclassified fractions show large variations among marker brackets and are sometimes as high as 20%, for example, around the marker *D17Mit81* and *D17Mit191*.

Table 2.4 Expected and estimated proportions of recombinant haplotypes (θ_9) in consecutive marker-brackets

Haplotype	Expected		Estimated θ_9			
	Distance[a] (cM)	θ_9[b]	*counting*[c]	*MLE*[d] (± s.e.)	*MLE* +*M*[e]	*M*[f]
D17Mit29-D17Mit198	0.90	0.039	0.179	0.110 (.012)	0.064	0.073
D17Mit198-D17Mit45	0.40	0.018	0.170	0.106 (.013)	0.000	0.141
D17Mit45-D17Mit81	0.50	0.022	0.282	0.179 (.017)	0.000	0.191
D17Mit81-D17Mit191	0.30	0.013	0.314	0.223 (.019)	0.109	0.191
D17Mit191-D17Mit175	0.50	0.022	0.201	0.119 (.013)	0.091	0.045
D17Mit175-D17Mit16	0.40	0.018	0.172	0.109 (.012)	0.047	0.090
D17Mit16-D17Mit214	0.60	0.026	0.178	0.106 (.012)	0.000	0.126
D17Mit214-D17Nds3	0.36	0.016	0.153	0.092 (.011)	0.013	0.106
D17Nds3-D17Mit148	1.33	0.056	0.185	0.113 (.012)	0.043	0.100
D17Mit148-D17Mit233	0.51	0.022	0.276	0.188 (.017)	0.079	0.169
D17Mit233-D17Mit11	1.05	0.045	0.286	0.211 (.019)	0.109	0.166

a) Map distance and markers order according to MGD (1999). b) Expected θ_9 using the formulae $\theta_t = [1-(1-\theta)^{t-2}(1-2\theta)]/2$ derived by Darvasi and Soller (1995) and based on MGD data. c) Counting method. d) Maximum likelihood estimation (MLE) method; standard error (s.e.) is estimated using the second derivative of the log of the likelihood function. e) MLE accounting for misclassification. f) Fraction misclassified.

Using two-point CRIMAP analysis results in a total length of 169.1 cM in generation nine for the region under study, assuming the same marker loci order as that of MGD. The multi-point analysis gives an order of marker loci that differs from MGD (1999) order in Table 2.2. The program could not significantly position maker loci D17Mit81 and D17Mit233. The positions of these two marker loci were forced to their most likely location. The new sequence of the marker loci is D17Mit198, D17Mit81, D17Mit29, D17Mit45, D17Mit214, D17Mit175, D17Mit191, D17Mit233, D17Nds3, D17Mit16, D17Mit148, and D17Mit11. The distances between consecutive markers are respectively 6.3, 4.1, 4.1, 9.9, 5.2, 4.1, 6.3, 7.5, 8.7, 8.7 and 17.8 cM giving a total length of 82.7 cM in F_9 and corresponding to $\theta_9 = 0.404$ between D17Mit198 and D17Mit11. Using the two-point analysis and assuming the new order of the marker loci, the distance between the two extreme markers is 187.3 cM. From the F_2, the estimated recombination fraction between *D17Mit45* and *D17Mit11* is 0.05 using the CRIMAP program; this result is similar to the estimate (0.046) obtained with MLE method.

Discussion

The generation of F_9 mice has been initiated with two inbred lines A/J and C57BL/6, and markers for which unique allele exists for each of the strains have been used to assess the genotype of mice. The allele frequencies are expected to be half, however, results in Table 2.1 show a significant deviation from the expected frequency for most of marker loci. According to Falconer and Mackay (1996), the expected standard deviation of the allele frequency in F_9 due to random drift is 0.0746. This indicates that the observed deviations from the expected frequency (0.5) can be explained by random drift and are not necessary due to systematic effects.

In the F_9 population, genotype frequency for marker-locus D17Mit191 departs from the expected frequency [Hardy-Weinberg Equilibrium (HWE)] (Table 2.2). The distribution of the test statistic is determined through simulation because we are dealing here with individuals among which relationships exist. Using the common chi-square table (with one degree of freedom) would be inappropriate in these circumstances. The target chromosomal segment contains the major histocompatibility complex that consists of genes regulating maternal-fetal interactions and resistance to diseases, therefore, the chromosomal segment could be responsible for a natural selection (Hedrick 1999; Snoek et al. 1998) in favor of heterozygous animals. In this case, one would expect deviation for closely linked markers. The most likely explanation for the observed deviation from Hardy-Weinberg of marker *D17Mit191* would be a misclassification. This conclusion is supported by the result in Table 2.5 showing a misclassified fraction of 0.191 for *D17Mit191* and *D17Mit81*.

The counting method is commonly used to estimate recombination fractions in crosses involving inbred lines (Blank et al. 1988; Heine et al. 1998). This method, however, cannot use information from individuals heterozygous at two marker loci in an AIL in absence of pedigree information because the gametic phase of these individuals cannot be determined. Using the expected genotype frequencies in Table 2.1 weighted by the counted number of recombinant haplotypes gives the expectation of the estimated proportion of recombinant haplotypes: $E(\hat{\theta}_t) = \frac{2(1-\theta_t)\theta_t + \theta_t^2}{2-(1-\theta_t)^2 - \theta_t^2}$. This shows that the counting method only results in unbiased estimates of θ_t if θ_t is equal to 0 or 0.5. For all other θ_t 's the counting method is biased upward which is also shown by the simulation study. The estimation of the proportion of recombinant haplotypes in an advanced intercross population using the counting method would result in unbiased estimates if

individuals from generation F_8 were backcrossed to one of the founder lines. The backcross would result in animals in which double heterozygotes are not present.

The maximum likelihood models presented in this paper use information of all individuals and can be used if genotyping information is only available from the target population (F_9) with no pedigree information. Results on proportion of recombinant haplotypes using MLE method are lower than those obtained from the counting method, but higher than would have been expected based on the MGD map distances (Table 2.4). The discrepancy between the expected and the estimated proportions of recombinant haplotypes may be due to differences between crosses used to map markers reported in the MGD data and the one used in the current experiment. Results from F_2 data, however, do not support this hypothesis. The MLE of recombination fraction (0.046) from an F_2 genotyping data on markers *D17mit45* and *D17mit11* gives an expected proportion of recombinant haplotype of 0.174 in F_9 according to Eq [1]. The F_9 data, in contrast, gives an estimate of 0.284 indicating that 63% more recombinants were accumulated in this particular chromosomal region than predicted by Eq [1]. Based on the simulation study, 10% misclassification at each locus will cut down θ_9 by 22%. This will bring the proportion of recombinant haplotypes to 0.221, which is still 27% higher than the expected value (0.174). Results from CRIMAP analysis give even stronger evidence than the MLE analysis for the high proportion of recombinant haplotypes. The estimated θ_9 is 0.3722, which still is twice as high as the prediction (0.1857). A possible explanation for this high proportion of recombinant haplotypes could be linked to natural selection favoring recombinants due to the presence of the major histocompatibility complex within the target chromosomal segment (Hedrick 1999; Snoek et al. 1998; Ritte et al. 1991; Hughes and Nei 1989).

As shown in Table 2.4, the maximum likelihood estimation method accounting for misclassifications gives estimates of the proportion of recombinant haplotypes that are closer to the expected values than results obtained using the MLE model. From the simulation results, however, it can be concluded that the proportion of recombinant haplotypes is underestimated when applying the MLE model that accounts for misclassification. Based on this model, the results should be corrected upward. The simulation shows also that the estimates of the misclassified fractions tend to be overestimated. The estimates of misclassified fractions should, therefore, be corrected downward.

Partly in this study, it is implicitly assumed that the MGD map distances are correct and used as reference. Our data set, according to the CRIMAP analysis, suggests however, a different order for the marker loci compared to that of the MGD (see result section). The change of order is expected given the markers density.

For the population used in this study, the observed deviations of allele frequencies from the expected frequency (0.5) can be explained by random drift and are not necessarily due to systematic effects. The counting method gives biased estimates of the proportions of recombinant haplotypes in F_9. This method, therefore, is not suitable for estimating the proportion of recombinant haplotypes in an advanced intercross population when pedigree information is not available. In such conditions, the maximum likelihood models developed in this study constitute alternatives for analyzing such data. Misclassification is an important factor to be considered when analyzing such data even though the accuracy of the estimates of misclassified fraction is low. The results of this study clearly show evidence of a strong accumulation of recombinant haplotypes within the target chromosomal region in the F_9, that is 63-100 % higher than the prediction applying Darvasi and Soller's theory. Results also suggest different markers' order compared to that reported by the mouse genome database (MGD 1999).

Acknowledgements

O.D.K acknowledges the International Livestock Research Institute (ILRI) for giving him training opportunity in laboratory analysis. He is also grateful to the leave provided by "Institut National des Recherches Agricoles du Bénin" to undertake this program. The authors are thankful to ILRI and the Netherlands Foundation for the Advancement of Tropical Research (WOTRO) for financial support.

References

Bailey DW (1971) Recombinant inbred strains. An aid to finding identity, linkage, and function of histocompatibility and other genes. *Transplantation* 11, 325-327

Blank RD, Campbell GR, Calabro A, d'Eustachio P (1988) A linkage map of mouse chromosome 12: localization of *Igh* and effects of sex and interference on recombination. *Genetics* 120, 1073-1083

Bovenhuis H, Van Arendonk JAM (1991) Estimation of milk protein gene frequencies in crossbred cattle by maximum likelihood method. *J Dairy Sci* 74, 2728-2736

Darvasi A, Soller M (1995) Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* 141, 1199-1207

Darvasi A (1997) Interval-specific congenic strains (ISCS): an experimental design for mapping a QTL into a 1-centimorgan interval. *Mamm Genome* 8, 163-167

Darvasi A (1998) Experimental strategies for the genetic dissection of complex-traits in animal models. *Nat Genet* 18, 19-24

Demant P, Hart AAM (1986) Recombinant congenic strains - a new tool for analyzing genetic traits determined by more than one gene. *Immunogenetics* 24, 416-422

Dietrich W, Katz H, Lincoln SE, Shin H-S, Friedman J, Dracopoli NC, Lander ES (1996) A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* 131, 423-447

Falconer DS, Mackay TFC (1996) *Introduction to quantitative genetics*. 4th edition, Longman, Edinburg, England

Green P, Fall K, Crooks S (1990) Documentation for CRIMAP Version 2.1

Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. *J Genet* 8, 299-309

Hedrick PW (1985) *Genetics of population*. Jones and Bartlett Publishers, Inc., Boston, USA

Hedrick PW (1999) Balancing selection and MHC. *Genetica* 104, 207-214

Heine D, Khambata S, Passmore HC (1998) High-resolution mapping and recombination interval analysis of mouse chromosome 17. *Mamm Genome* 9, 511-516

Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc Natl Acad Sci* 86, 958-962

Iraqi F, Clapcott SJ, Kumari P, Haley CS, Kemp SJ, Teale AJ (2000) Fine mapping of trypanosomiasis resistance loci in murine advanced intercross lines. *Mamm Genome* 11, 645-648

Kemp SJ, Iraqi F, Darvasi A, Soller M, Teale AJ (1997) Localization of genes controlling resistance to trypanosomiasis in mice. *Nat Genet* 16, 194-196

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685

Mouse Genome Database (MGD), Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, Maine. World Wide Web (URL:http://www.informatics.jax.org/). (July 9, 1999)

Nadeau JH, Singer JB, Matin A, Lander ES (2000) Analysing complex genetic traits with chromosome substitution strains. *Nat Genet* 24, 221-225

Nelder JA, Mead R (1965) A simplex method for function minimization. *Computer J* 7, 308-313

Ritte U, Neufeld E, O'hUigin C, Figeroa F, Klein J (1991) Origins of H-2 polymorphism in the house mouse: II. Characterization of a model population and evidence for heterozygous advantage. *Immunogenetics* 34, 164-173

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: A Laboratory Manual*. 2nd ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press)

Snoek M, Teuscher C, Van Vugt H (1998) Molecular analysis of the major MHC recombinational hot spot located within the *G7c* gene of the murine class III region that is involved in disease susceptibility. *J Immunol* 160, 266-272

Talbot CJ, Nicod A, Cherny SS, Fulker DW, Collins AC, Flint J (1999) High-resolution mapping of quantitative trait loci in outbred mice. *Nat Genet* 21, 305-308

CHAPTER III

A model for population growth of laboratory animals subjected to marker-assisted introgression: how many animals do we need?

Olorounto D. KOUDANDÉ[1,2], Peter C. THOMSON[3] & Johan A.M. Van ARENDONK[1]

[1] *Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands*
[2] *Institut National des Recherches Agricoles du Bénin, B.P. 884, Cotonou, République du Bénin*
[3] *Biometry Unit, Department of Crop Sciences, University of Sydney, NSW, 2006, Australia*

Published in Heredity 82 (1999) 16-24

Abstract

This study provides methods for calculating the mean and variance of the number of animals with the desired genotype in each backcross generation for a marker assisted introgression experiment. The ultimate goal is to produce animals which are homozygous for the desired loci. The methods have been developed specifically for experiments with inbred lines. The model assumes a Poisson distribution for litter size, and is similar to that used in stochastic versions of population dynamics models. Certain biological parameters must be specified as well as parameters under the control of the breeder. These methods can be utilised in designing an experiment to determine the number of founder animals required, given the number of animals required at the completion of the backcross process and vice versa. Consideration is given to minimizing the total amount of genotyping over the entire experiment, by varying the number of times each backcrossed male is used. In addition, an outline is given for an adaptive design that allows for changes in male usage to be made during the experiment.

Keywords: experimental design, inbred lines, MAI, Poisson distribution, population growth.

Introduction

Trypanosomiasis transmitted by the tsetse fly occurs in 37 countries in Africa, and there is evidence of genetic resistance to trypanosomiasis in some cattle breeds (Murray & Trail 1984). A program has commenced at the International Livestock Research Institute in Nairobi, Kenya, with the aim of identifying genes for trypanotolerance in N'Dama cattle (Teale 1993). The intention is to use marker assisted-introgression (MAI) to develop trypanotolerant cattle types. Trypanoresistance is well known in some laboratory mouse strains (Morrison et al. 1978), and three autosomal chromosomal regions have been identified (Kemp et al. 1997). Consequently, mice are being used as an animal model in a pilot study to assess the effectiveness of MAI of trypanotolerance genes for cattle.

In the design stage of the experiment, a literature survey was undertaken to identify how many founder donor animals would be needed for the experiment. Although there is a considerable volume of material dealing with MAI, relatively little deals with experimental design. Gama et al. (1992) studied introgression strategies of transgenes in pigs and considered the effect of various strategies (number of male and female donor animals) on the number to be genotyped to detect some deleterious transgene effect. Groen & Smith (1995) conducted a stochastic simulation study of the efficiency of marker-assisted introgression in livestock, and investigated the number of animals with the desired

genotype. Several studies have reported methods for reducing the number of backcross generations required for the introgression process (e.g. Hospital et al. 1992; Hillel et al. 1993; Yancovich et al. 1996). Hillel et al. (1990), using genomic selection based on DNA fingerprints for gene introgression in a breeding programme, investigated the theoretical distribution of the proportion of the donor's genome in successive backcross generations. Visscher et al. (1996) focussed more on the proportion of the recipient genotypes (background genotypes) existing after specified numbers of backcrossing generations. They studied the simultaneous effect of introgressing an allele and selecting for a desired genomic background. They also looked at a selection index combining markers and phenotypic information. However, none of these studies provides strategies for determining the number of animals required for the experiment. The determination of this number is quite important to plan properly the logistics necessary for such an experiment and to estimate its cost.

Consequently, the present paper intends to provide a mathematical model for estimating the mean and variance of the number of animals at each backcross generation resulting in the targeted number of favourable animals, i.e. those with the desired chromosomal region(s), at the end of the backcross process. The ultimate goal is to intercross those animals to produce sufficient animals that are homozygous for the desired loci in one breeding cycle if possible. Opportunities might exist for the breeder to repeat the intercross process from heterozygous individuals to obtain the required number of animals.

The objective of the paper is to predict the amount of genotyping required during a backcross experiment in order to produce a specified number of individuals with desired genotypes. In addition, the consequences of reducing the risk of not obtaining enough animals of the desired genotype to a set level are studied. Aspects considered here are: the number of founder animals; the number of times each male is used with the objective of minimizing the genotyping workload; the estimated number of animals to be genotyped for the whole experiment and at each generation, as well as the expected number of animals with a favourable genotype. The effect of the number of backcross generations, the number and the size of the regions to introgress as well as the influence of risk limitation will be studied.

Theory

Model of Population Growth

Some assumptions have been made: the donor and recipient lines used in the experiment are fully inbred; the introgressed alleles or regions are unambiguously identifiable in founder donor as well as in crossbred animals, and they lie on autosomal chromosomes. Each of a fixed number of males and females selected from the donor line is mated only once with an animal from the recipient line to produce F_1 offspring. The first backcross generation is produced by mating F_1 animals to animals of the recipient line. Selection of animals for the production of subsequent backcross generations is based solely on the presence of all the entire desired chromosomal regions from the donor strain. Animals which did not inherit all regions or animals carrying one or more recombinations in any of the regions are not selected. For the development of the model, it is also assumed that all the favourable animals will be used in subsequent production of the next generation; in practice this does not need to be the case though, depending on the number of animals available. Because of differences in reproductive capacities, favourable females are mated with only one male in the recipient line, whereas selected males can be mated to more than one recipient line female. The number of times that males are used will be treated as a design parameter. Attention will also be paid to the situation in which only males are genotyped and selected during the backcross generations. It will also be assumed that there are always sufficient recipient animals available for mating.

A further assumption is that litter size varies stochastically, and independently, from female to female. As a working approximation, a Poisson distribution for litter size is assumed (see e.g. Foulley et al. 1987; Matos et al. 1997), the major consequence being that the variance of litter size is taken as being equal to the mean.

In developing the theory, the mean and variance of the number of animals with the desired genotype at each generation are provided. Derivation of the full probability distribution of the number of animals at each generation is a more intractable problem. These issues are considered in mathematical theories of population dynamics (e.g. Renshaw, 1991). However, it is believed that use of summary statistics will characterise the population sufficiently well, and also allows a relatively simple development of the theory.

The following notation will be used here:

t = generation number (0 = F1; 1, 2, 3, ... = backcross generation 1, 2, 3, ...);

M_t = number of favourable males at backcross generation t;

F_t = number of favourable females at backcross generation t;

$N_t = M_t + F_t$, the total number of favourable animals at generation t;

n = number of males and number of females initially selected from each strain to start the introgression process (i.e. $2n$ donor animals and $2n$ recipient animals);

p = probability of inheriting the desired chromosomal region(s), i.e., favourable;

r = number of times each favourable male is used for breeding;

s = sex ratio in offspring at breeding age;

λ = mean number of animals produced in a litter and surviving to breeding age.

Although it would usually be assumed that the sex ratio s would equal ½, other values might occur as a result of differential mortality of males and females up to the time of breeding. Note that the probability of inheriting the desired chromosomal region(s) is a function of the number and the size of the region(s) to be introgressed. Let d_i be the length of a region (Morgans) and z the number of independent regions. With unequal chromosomal region lengths and applying Haldane's (1919) Poisson model for crossing-over, p is calculated as $p = \left(\frac{1}{2}\right)^z \prod_{i=1}^{z} e^{-d_i}$ which reduces to $p = \left(\frac{1}{2}\right)^z e^{-zd}$ when all regions are of equal length.

Assuming that the parameters r, p, s and λ are constant over generations, it may also be shown that there are explicit solutions for the mean and variance of N_t, namely $\mathrm{E}(N_t) = 2n\lambda\left[p\lambda(rs+1-s)\right]^t$ and $\mathrm{var}(N_t) = \mathrm{E}(N_t)\left[\beta + \gamma \mathrm{E}(N_t)\right]$ where

$$\beta = 1 - \frac{p\lambda(r^2 s + 1 - s)}{(rs+1-s)\left[p\lambda(rs+1-s)-1\right]}$$

and

$$\gamma = \frac{p(r^2 s + 1 - s)}{2n(rs+1-s)\left[p\lambda(rs+1-s)-1\right]}.$$

As an illustration (Fig. 3.1), 20 simulations were performed, each starting from $2n = 36$ donor animals with $r = 4$; $\lambda = 5$; $s = ½$; $p = 0.0686$ (based on $z = 3$ and $d = 20$ cM). This also shows the means ± 2 standard deviations, $\mathrm{E}(N_t) \pm 2\left[\mathrm{var}(N_t)\right]^{1/2}$. As expected, relatively few simulated values lie outside these limits.

Constant breeding strategy

Two parameters (n and r) are under control of the breeder. When specifying that a certain mean number (say N_{mean}) of favourable animals are to be available after k backcross generations and for a specified value of r, the initial number of animals required is clearly

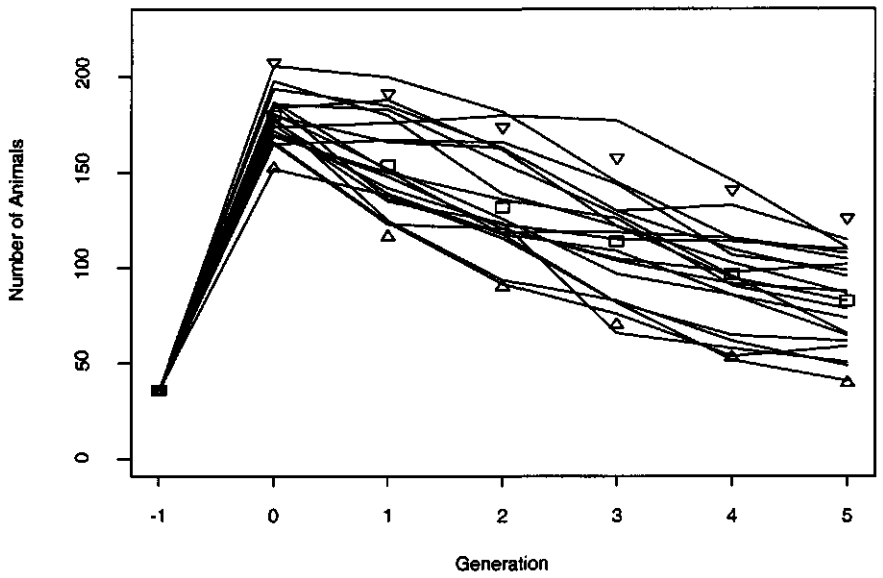

Fig. 3.1 Simulation of the number of favourable animals at each generation, N_t. Twenty simulations are shown all starting with $2n$ = 36 founder animals, other parameters are r=4, λ=5; s=½; p=0.0686 (based on z = 3 and d = 20 cM). Also shown are the theoretical mean (square) ± 2 standard deviations (triangle). Generation "-1" refers to the founders, "0" to the F1 generation, and subsequent generations are backcrosses.

$n = N_{mean} \Big/ \left\{ 2\lambda \left[p\lambda(rs+1-s) \right]^k \right\}$. For this strategy, however, there will be (approximately) a 50% risk of obtaining fewer animals than this at the end of the backcross process. An alternative strategy would be to specify the minimum number of favourable animals, N_{min}, and the nominated risk of obtaining fewer animals than this, α.

Provided that N_{min} is not too small, the distribution of the number of favourable animals at generation k can be taken as approximately normal, with mean $\mu = E(N_k)$ and variance $\sigma^2 = \text{var}(N_k)$. So for a chosen value of r, we select the value of n so that $P(N_k < N_{min}) \leq \alpha$. This may be approximated as the value n satisfying

$$\frac{N_{\min} - E(N_k)}{\sqrt{\text{var}(N_k)}} \leq \Phi^{-1}(\alpha)$$

where $\Phi^{-1}(\alpha)$ is the 100α percentage point of the standard normal distribution. However, when N_{min} is small, a better approximation to normality is achieved by using a square-root transformation. These means and variances may be approximated as

$$\mathrm{E}\left(\sqrt{N_k}\right) \approx \sqrt{\mathrm{E}(N_k)} + \tfrac{1}{4}\,\mathrm{var}(N_k)[\mathrm{E}(N_k)]^{-3/2} \quad \text{and} \quad \mathrm{var}\left(\sqrt{N_k}\right) \approx \tfrac{1}{4}\,\mathrm{var}(N_k)/\mathrm{E}(N_k),$$

and n then chosen to satisfy

$$\frac{\sqrt{N_{\min}} - \mathrm{E}\left(\sqrt{N_k}\right)}{\sqrt{\mathrm{var}\left(\sqrt{N_k}\right)}} \leq \Phi^{-1}(\alpha).$$

Note that different values of male usage, r, result in different minimum values of n. One possible option is to choose the value of r that would require the smallest expected number of animals to be genotyped during the entire experiment (amount of genotyping). Because for any backcross generation, the amount of genotyping required has a mean of $p^{-1}\mathrm{E}(N_t)$ if males and females are genotyped, the total expected amount of genotyping over k backcross generations would be

$$EG_k = p^{-1}\sum_{t=1}^{k}\mathrm{E}(N_t) = 2n\lambda^2(rs+1-s)\frac{[p\lambda(rs+1-s)]^k - 1}{p\lambda(rs+1-s)-1}.$$

Instead of genotyping males and females, one could only genotype and select males from generation 1 to $k-1$, and subsequently males and females in generation k, which results in a total expected amount of genotyping of:

$$EG_k = p^{-1}\left\{\mathrm{E}(N_k) + \sum_{t=1}^{k-1}\mathrm{E}(M_t)\right\} = 2n\lambda^2 sr\left[(\lambda spr)^{k-1} + s\frac{\left[(\lambda spr)^{k-1} - 1\right]}{\lambda spr - 1}\right].$$

Alternatively, a constant amount of genotyping each generation may be the preferred option. In this situation, we need the growth rate to be (close to) one; this is achieved by choosing r to be the nearest integer to $(p\lambda s)^{-1} - s^{-1} + 1$, that is $2/(p\lambda)$ - 1 assuming s = ½.

Adaptive breeding strategies

Another issue is that the breeding strategy may be adaptive in the sense that it might be changed between generation 0 and k, or perhaps extended for additional generations. Although n cannot be altered, r may be altered during the programme, particularly when N_t at generation t becomes critically small. For this, we need to calculate the conditional mean and variance of N_k given the actual number of males and females at generation t. These may be calculated as $\mathrm{E}(N_k \mid M_t, F_t) = p\lambda(rM_t + F_t)[p\lambda(rs+1-s)]^{k-t-1}$ and $\mathrm{var}(N_k \mid M_t, F_t) = \mathrm{E}(N_k \mid M_t, F_t)[\beta + \gamma\mathrm{E}(N_k \mid M_t, F_t)]$, where β and γ are as defined above. So at each intervening generation, t = 0, ..., k-1, we may select the value of r so that

$$\frac{N_{\min} - \mathrm{E}\left(N_k \mid M_t, F_t\right)}{\sqrt{\mathrm{var}\left(N_k \mid M_t, F_t\right)}} \leq \Phi^{-1}(\alpha).$$

Another adaptation to face situations in which the number of favourable animals has dropped below the desired level is to produce more offspring than planned in subsequent generations.

Numerical application

In this section, consequences of the theory outlined previously will be investigated numerically. For all calculations, regions to be introgressed are assumed to be of equal size. The sex ratio (s) was assumed to be ½ and the mean litter size (λ) assumed to be 5. First the situation in which both males and females are genotyped will be considered. Subsequently, results for situations in which only males are genotyped during the backcrossing process are presented.

Table 3.1 Expected number of animals to be genotyped (EG_k) for breeding scheme (r), the expected number (N_{mean}) of favourable animals required at the end of backcrosses being 50. (p = probability of inheriting the desired chromosomal region(s))

Back-crosses	No. of regions	Size of region (cM)	p	r 1	2	3	4	5
5	1	0	0.5	165 (.051)	-	-	-	-
		10	0.4524	195 (.085)	-	-	-	-
		20	0.4094	232 (.139)	-	-	-	-
	2	0	0.25	672 (1.63)	410 (.215)	-	-	-
		10	0.2047	1166 (4.45)	618 (.586)	-	-	-
		20	0.1676	2191 (12.09)	994 (1.59)	684 (.378)	-	-
	3	0	0.125	6324 (52.42)	2285 (6.90)	1345 (1.63)	992 (.536)	-
		10	0.0926	21415 (234.99)	6370 (30.94)	3167 (7.34)	2058 (2.40)	1555 (.967)
		20	0.0686	79770 (1053.17)	20652 (138.68)	8889 (32.91)	5074 (10.78)	3445 (4.33)
10	1	0	0.5	166 (.000)	-	-	-	-
		10	0.4524	198 (.001)	-	-	-	-
		20	0.4094	239 (.003)	-	-	-	-
	2	0	0.25	893 (.536)	-	-	-	-
		10	0.2047	2205 (3.96)	691 (.068)	-	-	-
		20	0.1676	7493 (29.27)	1311 (.507)	-	-	-
	3	0	0.125	72634 (549.75)	5440 (9.53)	1785 (.536)	-	-
		10	0.0926	1027920 (11044.8)	45793 (191.53)	7819 (10.78)	3049 (1.15)	1856 (.187)
		20	0.0686	16881990 (221834)	593469 (3846.72)	67399 (216.63)	16018 (23.26)	6430 (3.75)

The numbers in parentheses are n, i.e. n donor males and n donor females are mated to n recipient females and n recipient males to produce the F_1 generation.

Genotyping males and females

In the first investigation, the mean number of favourable animals was specified (N_{mean} = 50), and the initial number of donor animals (n males and n females) determined (Table 3.1). Also shown is the expected number of animals to be genotyped for the whole experiment (EG_k). The probability p of inheriting the desired genotype decreases when z or

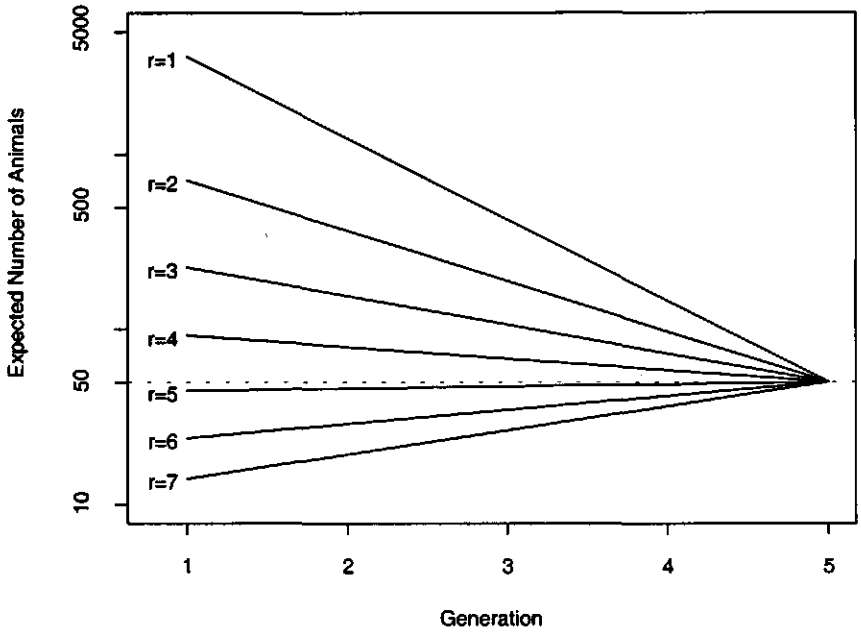

Fig. 3.2 Expected number of favourable animals ($E(N_t)$) at each generation of backcross for different breeding schemes (r). Fixed parameters were N_{mean}=50 at k=5 backcross generations,; λ=5; s=½; p=0.0686 (based on z = 3 and d = 20 cM). The broken line is drawn at N_{mean}=50.

d increases. Note that in some situations, values of n less than one result in the required number of animals in the target generation. In these situations, entries for higher values of r have been left empty, as higher usage of males is clearly not necessary. Both n and the number of animals to be genotyped for the whole experiment (EG_k) decrease with increasing r and p, but increase when the number of backcrosses k becomes higher. Figure 3.2 shows the pattern of $E(N_t)$ over the backcross generations which converges to a fixed value of N_{mean} = 50. Note that a plot of the expected amount of genotyping over time would show the same pattern, converging to 729 animals after k = 5 backcross generations.

Table 3.2 Expected number of animals to be genotyped (EG_k) for breeding scheme (r), the minimum number (N_{min}) of favourable animals required at the end of backcrosses being 50, α = 0.05. (p = probability of inheriting the desired chromosomal region(s))

Back-crosses	No. of regions	Size of region (cM)	p	r = 1	2	3	4	5
5	1	0	0.5	1761 (.546)	-	-	-	-
		10	0.4524	1474 (.639)	-	-	-	-
		20	0.4094	1293 (.775)	-	-	-	-
	2	0	0.25	1440 (3.51)	1927 (1.01)	4247 (.659)	-	-
		10	0.2047	2081 (7.94)	1869 (1.77)	3028 (.907)	-	-
		20	0.1676	3482 (19.22)	2223 (3.56)	2628 (1.45)	4153 (.924)	-
	3	0	0.125	9122 (75.62)	3935 (11.89)	3115 (3.79)	3436 (1.86)	4584 (1.20)
		10	0.0926	29332 (321.87)	9578 (46.53)	5556 (12.88)	4483 (5.24)	4462 (2.77)
		20	0.0686	106157 (1401.56)	28966 (194.51)	13535 (50.11)	8652 (18.38)	6798 (8.55)

The numbers in parentheses are n, i.e. n donor males and n donor females are mated to n recipient females and n recipient males to produce the F_1 generation.

For the second investigation, the minimum number of animals after k backcross generations was specified (N_{min} = 50), with a risk of α = 0.05 of obtaining fewer than this number (Table 3.2). This results in an increase of all corresponding values of EG_k and n compared with the values in Table 3.1 where the mean was set at 50; similarly n decreases with increasing p, increasing r and for greater number of backcross generations. The expected amount of genotyping (EG_k) behaves differently however; for smaller values of p, EG_k initially decreases with increasing r, reaches a minimum value and then starts increasing (Table 3.2). However, the initial decrease is not observed when p exceeds 0.216 (for the given values of λ, s).

Table 3.3 Means and variances of N_5 for α = 0.50 and α = 0.05; z = 3 and d = 20 cM

	α = 0.50		α = 0.05	
r	E(N_5)	var(N_5)	E(N_5)	var(N_5)
1	50	76	67	101
2	50	107	70	150
3	50	166	76	252
4	50	269	85	460
5	50	444	99	876
6	50	725	118	1713
7	50	1163	146	3389

Table 3.4 Expected number of animals to be genotyped (EG_k) for r = 4, k = 5, z = 3 and d = 20 cM

	N_{min}		
α	25	50	100
0.50†	2537	5074	10148
0.10	4561	7705	13650
0.05	5351	8652	14833

†N_{min} corresponds to N_{mean} when α=0.50

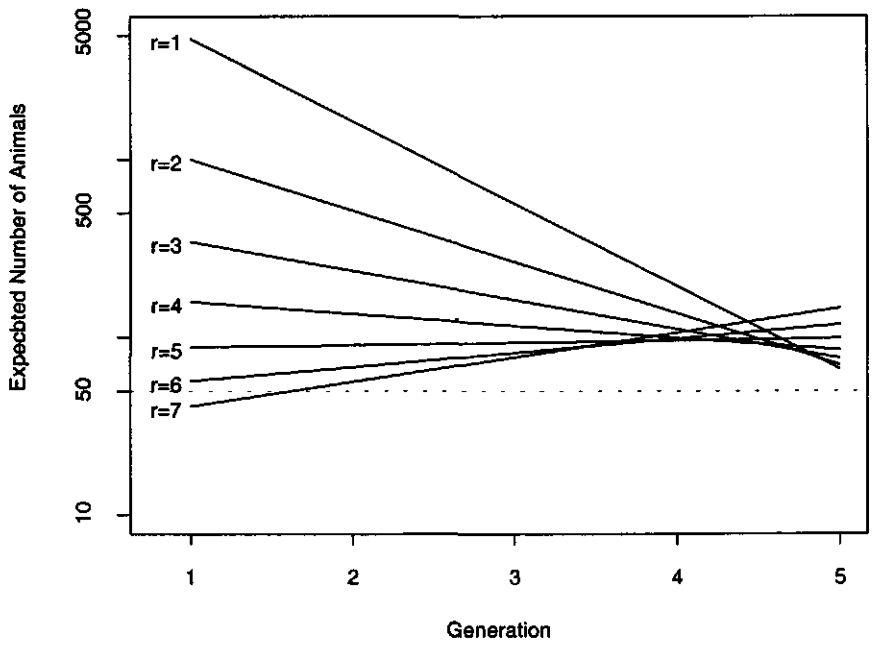

Fig. 3.3 Expected number of favourable animals ($E(N_t)$) at each generation of backcross for different breeding schemes (r). Fixed parameters were N_{min}=50 at k=5 backcross generations with risk α=0.05; λ=5; s=½; p=0.0686 (based on z = 3 and d = 20 cM). The broken line is drawn at N_{min}=50.

Figure 3.3 shows the behaviour of $E(N_t)$ over the backcross generations. The mean population trajectories do not converge to the same value for all values of r; Table 3.3 shows the means and variances of N_t at the fifth generation for a given N_{mean} or N_{min}. With high r-values, the initial number of founders ($2n$) is relatively small, but the final mean ($E(N_k)$) tends to be higher, compared with the designs with a lower r-value. When the number of founders is small, there is relatively more variability in the number of favourable animals in F_1 and subsequent backcross generations. Consequently, the mean in the target backcross generation, $E(N_k)$, needs to be relatively high in order to 'ensure' that the minimum number of favourable animals is achieved.

As expected in the third investigation (Table 3.4), increasing the number of animals required results in an increased amount of genotyping, as does reducing the risk level. Further, doubling the number of animals required doubles the amount of genotyping when N_{mean} is specified. However, when the risk in considered (i.e. for α<0.50), doubling N_{min}, leads to a less than doubling of the amount of genotyping.

Genotyping males only

Males in general have a higher reproductive rate than females and consequently it might be interesting to look at the situation in which only males are genotyped and selected. During the backcrossing generations these selected males can be mated to ungenotyped females from the recipient line. Males and females in the final backcross generation need to be genotyped and selected for the intercross. Table 3.5 gives the number of founder animals n that are needed in order to have a risk of $\alpha = 0.05$ of obtaining fewer than $N_{min} = 50$ animals after five generations of backcrossing for the alternative of genotyping all animals ($r_F = 1$) or males only ($r_F = 0$). For higher values of r (≥ 4) the expected amount of genotyping is smaller by genotyping males only. However, if there is a limit to r then the Table indicates there are situations where genotyping females would be optimal with respect to the amount of genotyping. Genotyping males only results in an increase in the number of animals being produced in the different generations as illustrated by the number of founder animals n in Table 3.5.

Table 3.5 Expected number of animals to start the experiment (n) and expected number of animals to be genotyped (EG_k) for a situation with ($r_F = 1$) and without ($r_F = 0$) genotyping and selection of females for $\alpha = 0.05$, $N_{min} = 50$, $k = 5$, $z = 3$ and $d = 20$ cM

	$r_F = 1$		$r_F = 0$	
r	n	EG_k	n	EG_k
1	1,401.41	106147	42075.30	635115
2	194.50	28963	1403.97	53656
3	50.11	13534	200.54	15458
4	18.39	8652	52.79	7714
5	8.55	6798	19.73	5309
6	4.73	6172	9.33	4492

Discussion

The model developed in this study can be used to assist the experimentalist in designing an MAI experiment. It provides important insights into parameters that need to be considered and the scale at which the introgression needs to be carried out in order to be successful. When the aim is to minimize the amount of genotyping there might be options to reduce this amount by using an adaptive breeding scheme when r is lower than the maximum reproductive rate of males. Using low levels of r during the initial generations and high values in later generations will lead to smaller values of EG_k than those given in Tables 3.1 and 3.2. The results obtained for a constant breeding scheme will serve as a good starting

point for the calculations in that case. An additional reason for a departure from a constant breeding strategy might be if the breeder wishes to introgress two regions and exclude a third region. Once the breeder has eliminated that third region, p would be changed to the "original" level for two regions. This situation can be accommodated by allowing p to vary over the generations using this adaptive approach. Setting the mean (N_{mean}) at 50, one would expect a number of progeny of 125 after one generation of intercrossing. Introgressing $z = 3$ regions each of length 20 cM, then $\left(\frac{1}{2}e^{-0.2}\right)^6 = 0.47\%$ (corresponding to 0.6 animal) of the offspring is expected to be homozygous for the three-locus genotype. This demonstrates that nominating an expected number of 50 animals at the end of the backcross generations is too low. Now consider the case where the minimum (N_{min}) is set at 50, and assume that the breeder used each male $r = 7$ times during the backcrossing process (Table 3.3). This would result in approximately 1.7 animals, still a low number. Nevertheless, opportunities might exist for the breeder to repeat the intercross process, or additionally to intercross any heterozygous animals resulting from the intercross.

During the experiment, the experimenter will try to modify the scheme when the results would permit this. Genotyping of animals will be stopped as soon as sufficient animals with a desirable genotype have been found. It is clearly demonstrated that the amount of genotyping can be reduced by using such an adaptive breeding scheme (Table 3.5). There might be practical limitations in implementing a scheme which results in minimum genotyping. To make efficient use of genotyping facilities, animals are genotyped in batches rather than individually. Because of culling of surplus animals at an early age, it may be difficult to go back to genotyping more males or females if insufficient males with favourable genotypes are found. These practical constraints need to be taken into account in executing a MAI programme in order to make efficient use of experimental and genotyping facilities.

The number of founder animals from the donor as well as the recipient strain (n female and n male) are represented as a decimal number (Tables 3.1 and 3.3). In practice this cannot be realised. Imposing a minimum value of 1 for n would under a constant breeding strategy lead to too large numbers at the end of the experiment. Values smaller than 1 indicate that during the initial generations of the backcrossing process the number of selected animals should be kept constant instead of increased, perhaps by not using all the available favourable animals.

In MAI the major aim of backcrosses is the recovery of the recipient genome (Hospital et al. 1992; Hillel et al. 1993; Yancovich et al. 1996). The number of generations of backcrossing is determined by the desired proportion of genes coming from the donor line for other parts of the genome. With five backcrosses, the expected proportion of the donor

genome in the recipient animal is still 1.562% and it is 0.049% after 10 backcross generations (Soller & Plotkin-Hazan 1977; Hillel et al. 1990). In this paper, selection is entirely based on the genotypic information for the identified regions. To reduce the number of backcross generations, genomic selection against the donor genes from other parts of the genome may be applied (Hospital et al. 1992; Hillel et al. 1993; Yancovich et al. 1996). Introgressing alleles simultaneously with genomic selection on recipient genotype (Visscher et al. 1996) would be another way of reducing backcross generations. These techniques accelerate the recovery of the recipient genome but will lead to an increase in the number of animals to be produced within a generation and the amount of genotyping per animal; indeed the genomic selection is applied to those individuals known to carry the relevant segments.

One of the goals of this study is to minimize the number of animals to be genotyped throughout the experiment. If we ignore the risk of obtaining too few animals in the target backcross generation, then Table 3.1 indicates that we need to select a strategy with a small n and large r. In fact the expected amount of genotyping would be minimized further by a low constant population size up to the second last generation (k - 1), then the highest possible increase in the last generation (i.e. high r). However, when risk is considered, this may no longer be the optimal strategy. As is seen in Table 3.2, the expected amount of genotyping is minimized by selecting a low or intermediate value of r; higher values of r can lead to an increase in the amount of genotyping involved. An option to minimise the amount of genotyping would be to only type, select and use favourable males during the backcross process. From Table 3.5 it follows that when the reproductive rate of males is sufficiently large, the expected total amount of genotyping can be reduced by genotyping males only, at the expense of producing a larger number of individuals. In designing an experiment, the amount of genotyping needs to be balanced against the number of experimental animals needed. One will first determine the number of backcross generations that is needed to obtain the acceptable proportion of donor genome in the newly formed line. To determine the optimum choice for the other parameters one will need to incorporate the costs of producing and keeping an animal, the genotyping costs, the costs of other resources and the time required to achieve the target.

Alternatively, the choice of "optimum" may also be driven by the genotyping capacity of the laboratory. Low values of r are usually associated with high rates of population decrease, resulting in a large volume of genotyping being conducted in the early generations (Figs. 3.2 and 3.3). Conversely, when risk is considered, a high value of r may lead to an increased amount of genotyping in the later generations (Fig. 3.3). Clearly, for the parameters nominated for Fig. 3.3, a choice of $r = 5$ would be optimal in terms of approximate constant genotyping over all backcross generations.

This model was developed to assist in the design stage of a marker assisted introgression experiment of trypanoresistance genes in mice. The findings are now to be exploited in this programme. More generally, this model when implemented in a computer spreadsheet programme, allows the breeder to assess the effect of changes in biological and design parameters on the population growth, as well as effects of random causes. Such a system could be incorporated into a computer-aided decision-support system for use by breeders. A spreadsheet program (Microsoft Excel) for this model is available from the corresponding author.

Acknowledgements

O.D.K. is grateful to WOTRO (Netherlands Foundation for the Advancement of Tropical Research) and ILRI (International Livestock Research Institute) for funding this PhD project. He also acknowledges the leave provided by INRAB (Institut National des Recherches Agricoles du Bénin) to undertake the research. P.C.T. acknowledges the study leave provided by the University of Sydney to undertake this research at Wageningen Agricultural University. The authors would also like to thank Luc Janss, Alan Teale and two anonymous reviewers for helpful comments on an earlier version of the manuscript.

References

Foulley JL, Gianola D, Im S (1987) Genetic evaluation of traits distributed as Poisson-binomial with reference to reproductive characters. *Theor Appl Genet* 73, 870-877

Gama LT, Smith C, Gibson JP (1992) Transgene effects, introgression strategies and testing schemes in pigs. *Anim Prod* 54, 427-440

Groen AF, Smith C (1995) A stochastic simulation study of the efficiency of marker-assisted introgression in livestock. *J Anim Breed Genet* 112, 161-170

Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. *J Genet* 8, 299-309

Hillel J, Schaap T, Haberfeld A, Jeffreys AJ, Plotzky Y, Cahaner A, Lavi U (1990) DNA fingerprints applied to gene introgression in breeding programs. *Genetics* 124, 783-789

Hillel J, Verrinder Gibbins AM, Etches RJ, Shaver DMcQ (1993) Strategies for the rapid introgression of a specific gene modification into a commercial poultry flock from a single carrier. *Poultry Sci* 72, 1197-1211

Hospital F, Chevalet C, Mulsant P (1992) Using markers in gene introgression breeding programs. *Genetics* 132, 1199-1210

Kemp SJ, Iraqi F, Darvasi A, Soller M, Teale AJ (1997) Localization of genes controlling resistance to trypanosomiasis in mice. *Nat Genet* 16, 194-196

Matos CAP, Thomas DL, Gianola D, Tempelman RJ, Young LD (1997) Genetic analysis of discrete reproductive traits in sheep using linear and nonlinear models: I. Estimation of genetic parameters. *J Anim Sci* 75, 76-87

Mood AM, Graybill FA, Boes DC (1974) *Introduction to the Theory of Statistics*. McGraw-Hill, New York

Morrison WI, Roelants GE, Mayor-Withey KS, Murray WI (1978) Susceptibility of inbred strains of mice to *Trypanosoma congolense*: correlation with changes in spleen lymphocyte populations. *Clin Exp Immunol* 32, 25-40

Murray M, Trail JCM (1984) Genetic resistance to animal trypanosomiasis in Africa. *Prev Vet Med* 2, 541-551

Renshaw E (1991) *Modelling Biological Populations in Space and Time*. Cambridge University Press, Cambridge

Soller M, Plotkin-Hazan J (1977) The use marker alleles for the introgression of linked quantitative alleles. *Theor Appl Genet* 51, 133-137

Teale AJ (1993) Improving control of livestock diseases. Animal biotechnology in the Consultative Group for International Agricultural Research. *Bioscience* 43, 475-483

Visscher PA, Haley CS, Thompson R (1996) Marker-assisted introgression in backcross breeding programs. *Genetics* 144, 1923-1932

Yancovich A, Levin I, Cahanar A, Hillel J (1996) Introgression of the avian naked neck gene assisted by DNA fingerprints. *Anim Genet* 27, 149-155

Appendix: Derivation of means, variances, and covariances

F_1 generation (t = 0)

Because there are $2n$ donor animals mated with $2n$ recipient animals, each mating giving rise to a litter with mean size λ, all of which are favourable, the mean number of male and female favourable animals, respectively, will be $E(M_0) = 2ns\lambda$ and $E(F_0) = 2n(1-s)\lambda$. Also, $var(M_0) = E(M_0)$ and $var(F_0) = E(F_0)$ from the Poisson litter size assumption, and $cov(M_0,F_0) = 0$ because of the independence (for a given n). Clearly, $E(N_0) = var(N_0) = 2n\lambda$.

Backcross generations ($t = 1,2,\ldots$)

Because the M_{t-1} males are used r times and the F_{t-1} females only once, there will be $rM_{t-1}+F_{t-1}$ litters resulting from matings with the recipient line, each litter with a mean and variance of λ. However, there is now a probability p of an offspring being favourable, so the conditional moments are

$$E(M_t|M_{t-1},F_{t-1}) = ps\lambda(rM_{t-1}+F_{t-1})$$

and

$$E(F_t|M_{t-1},F_{t-1}) = p(1-s)\lambda(rM_{t-1}+F_{t-1}),$$

with

$$var(M_t|M_{t-1},F_{t-1}) = E(M_t|M_{t-1},F_{t-1},),$$

$$var(F_t|M_{t-1},F_{t-1}) = E(F_t|M_{t-1},F_{t-1},),$$

and

$$cov(M_t,F_t|M_{t-1},F_{t-1}) = 0.$$

The marginal (unconditional) moments are obtained using standard results (see e.g. Mood *et al.*, 1974, p. 157), namely, means of

$$E(M_t) = ps\lambda\times[rE(M_{t-1})+E(F_{t-1})],$$

$$E(F_t) = p(1-s)\lambda\times[rE(M_{t-1})+E(F_{t-1})],$$

and variances

$$var(M_t) = (ps\lambda)^2\times [r^2var(M_{t-1})+var(F_{t-1})+2rcov(M_{t-1},F_{t-1})] + ps\lambda\times[rE(M_{t-1})+E(F_{t-1})],$$

$$var(F_t) = (p(1-s)\lambda)^2\times[r^2var(M_{t-1})+var(F_{t-1})+2rcov(M_{t-1},F_{t-1})] + p(1-s)\lambda\times[rE(M_{t-1})+E(F_{t-1})],$$

and covariance

$$cov(M_t,F_t) = (p\lambda)^2s(1-s)\times[r^2var(M_{t-1})+var(F_{t-1})+2rcov(M_{t-1},F_{t-1})].$$

Also,

$$E(N_t) = p\lambda\times[rE(M_{t-1})+E(F_{t-1})],$$

and

$$\mathrm{var}(N_t) = (p\lambda)^2\times[r^2\mathrm{var}(M_{t-1})+\mathrm{var}(F_{t-1})+2r\mathrm{cov}(M_{t-1},F_{t-1})] + p\lambda\times[r\mathrm{E}(M_{t-1})+\mathrm{E}(F_{t-1})].$$

These means, variances, and covariance for M_t, F_t, and hence N_t, can be obtained by calculating the moments recursively. Alternatively, solutions for these may be obtained explicitly, as reported in the main text. These results may be proven by induction.

CHAPTER IV

Strategies to optimize marker-assisted introgression of multiple unlinked QTL

Olorounto D. KOUDANDÉ[1], Fuad IRAQI[2], Peter C. THOMSON[3], Alan J. TEALE[4], and Johan A. M. Van ARENDONK[1]

[1] *Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands*
[2] *International Livestock Research Institute (ILRI), P.O. Box 30709 Nairobi, Kenya*
[3] *Biometry Unit, Department of Crop Sciences, University of Sydney, NSW, 2006, Australia*
[4] *Institute of Aquaculture, University of Stirling, Stirling, Scotland FK9 4LA, UK*

Published in Mammalian Genome 11 (2000) 145-150

Abstract

To optimize designs to implement marker-assisted introgression programs aiming to introgress three unlinked quantitative trait loci (QTL), the present paper studies different alternatives versus a traditional backcross or intercross phase. Four alternative backcross strategies appear to be more advantageous by having 50% less genotyping load than a traditional backcross strategy tracking all three QTL at a time through a single line. A multiplication phase following the selection of homozygous animals at the three QTL as an intercross alternative allows doubling of the number of homozygous animals in a mouse model compared with the first intercross generation. Within the same model, a second intercross alternative with individuals carrying all three QTL at the first intercross results in a 12-fold increase in the number of homozygous animals obtained in the first intercross generation. The same ranges of decrease are observed in the number of animals to be genotyped and the number of genotypings when targeting a fixed number of homozygous animals. An option, with two lines each carrying two QTL through the backcross phase and coupled with the second intercross alternative, appears to be the best introgression alternative. This option requires 76% fewer genotypings, 68% fewer animals to be genotyped, and costs 75% less than an option in which all three QTL are introgressed through a single line.

Introduction

The discovery of DNA microsatellites that are highly polymorphic has stimulated research on detection and mapping of QTL in animal species. Many studies have been undertaken using DNA microsatellites (genetic markers) in the selection of animals in breeding programs, a process known as marker-assisted selection (e.g., Kashi et al. 1990; Lande and Thompson 1990). The use of those DNA microsatellites to introgress genes from a donor to a recipient animal, process known as marker-assisted introgression (MAI), is under increasing consideration for the purpose of livestock improvement. Most studies of MAI have focused on introgression of a single QTL. Nonetheless, valuable information has been provided by these studies, such as the number of animals with the desired genotype needed during the backcross phase (Gama et al. 1992; Groen and Smith 1995). Methods to reduce the number of backcross generations required for introgressing a single QTL have also been proposed (Visscher et al. 1996; Yancovich et al. 1996; Markel et al. 1997; Wakeland et al. 1997).

Most characteristics of animals are, however, under multigene control, and breeders may want to simultaneously select for or introgress more than one QTL. Recently, two

studies on multiple QTL introgression have been published (Hospital and Charcosset 1997; Koudandé et al. 1999). The first deals with the number of QTL introgressed at the same time and the uncertainty of the position of the QTL within a marked chromosomal segment. Throughout, the study focuses on deriving the minimum number of individuals to be genotyped to obtain at least one individual with the desired marker genotypes at all QTL at the end of the backcross program. The second study is based on the three trypanosomiasis-resistance QTL discovered in mouse (Kemp et al. 1997) and focuses on the number of animals to be genotyped in the backcrossing phase of an introgression scheme. It shows that this number becomes extremely large when selecting all three QTL throughout the backcross process in a single line. To minimize the number of animals to be genotyped, Markel et al. (1997) and Koudandé et al. (1999) suggest selecting only backcross males until the second-last generation of the backcross phase and to increase their use. So far, little attention has been paid to the optimization of the intercrossing phase.

Introgressing more than one QTL, rather than focusing on a single QTL or gene, leads to an increase in the required number of animals. The present paper, following that of Koudandé et al. (1999), aims at optimizing the introgression program that is targeting three unlinked QTL. Optimization aims at decreasing the total number of animals to be genotyped, given a specified number of individuals homozygous at three QTL at the end of the introgression program. First we examine backcross and intercross phases separately and develop some alternatives for each phase. Second, we study combinations of backcross and intercross alternatives. Introgression strategies are compared with respect to the number of animals to be genotyped, the number of genotypings, and the costs.

Materials and methods

A traditional introgression program consists of the production of an F_1 generation from the founder animals, followed by a number of backcrossing generations aiming to reduce the proportion of the donor genome (Gama et al. 1992; Groen and Smith 1995). Subsequently, an intercrossing phase is used to fix the introgressed alleles.

In the present study aiming to introgress three QTL, we use the backcross design, consisting in genotyping and selecting only males from F_1 up to the second-last generation of backcross (Markel et al. 1997; Koudandé et al. 1999). Further, it is assumed that the donor and recipient lines used in the experiment are fully inbred, that all chromosomal regions to be introgressed are of equal size d (in Morgan) and can be accurately identified through genetic markers in donor as well as in crossbred animals. These chromosomal regions contain the QTL of interest and will be referred to as QTL or QTL regions.

The following notations will be used throughout the paper:

p = probability of inheriting one non-recombinant QTL region from a heterozygous parent, $p = \left(\frac{1}{2}\right)e^{-d}$ where e is the base of the natural logarithm and d the length of the chromosomal region to be introgressed expressed in Morgan (Koudandé et al. 1999),

s = probability that an offspring is male,

λ = mean number of offspring per female surviving till breeding age,

z = number of independent QTL regions ($z = 3$ unless specified otherwise),

n = number of males and number of females initially selected from each strain or breed to start the introgression process (for example, $2n$ donor animals and $2n$ recipient animals to produce F_1 generation),

r = number of females mated with one male during the backcross phase,

k = number of backcross generations,

t = generation number: $t = 1, 2, 3\ldots$

In the following sections, we will describe the backcross as well as the intercross phase, starting in each case with the traditional (or basic) strategy, and followed by alternatives. Subsequently, we will specify the combinations of backcross and intercross alternatives, that is, introgression programs that have been evaluated.

1. Backcross phase

1.1. Traditional backcross (**tB**)

Traditional backcross is based on selection of animals carrying the three QTL regions throughout the backcross generations. At the end of the backcross, individuals carrying the three QTL regions of interest are used for the intercross. This traditional backcross is described in more detail by Koudandé and colleagues (1999) and consists in moving three QTL regions from one breed to another through a single line, that is, one population in which selected parents for the next generation are heterozygous at all three QTL.

From backcross generation 1 up to generation k-1, the expected number (M_t) of animals heterozygous at three QTL in generation t is $M_t = 2n\lambda s\left(\lambda s p^3 r\right)^t$. At generation k, males and females must be selected to start the intercross phase. The expected number (N_0) of animals (males and females) heterozygous at all three QTL is:

$$N_0 = 2n\lambda\left(\lambda s p^3 r\right)^k \qquad [1]$$

1.2. Alternative backcross strategies (**aB**)

Four alternative backcross strategies are successively described. These alternatives are based on selection of individuals carrying one or two QTL regions during the backcross. Depending on the alternative, selection of individuals results in the formation of two or three different backcross lines. These lines are subsequently crossed to produce animals carrying the three QTL regions.

First alternative (**aB**$_1$). In alternative **aB**$_1$, animals heterozygous at a single QTL region are selected, resulting in three different lines, each corresponding to one QTL. After the last backcross generation, two lines are crossed while the third is intercrossed. Females heterozygous at two QTL are selected from the cross, while males homozygous at the other QTL are selected from individuals resulting from the intercross. Selected animals are mated, and from resulting offspring, individuals heterozygous at all three QTL regions are selected to start the intercrossing phase. From backcross generation 1 up to generation k-1, the expected number (M_t) of animals heterozygous at a single QTL for generation t within each line is $M_t = 2n\lambda sp^3(\lambda sr)^t p^{t-1}$. At generation k, the expected number (N') of animals (males and females) heterozygous at one QTL for each line is $N' = 2n\lambda p^3(\lambda sr)^k p^{k-1}$. After the two generations of crosses between lines, the expected number (N_1) of animals heterozygous at all three QTL is:

$$N_1 = 4n\lambda^3(1-s)^2 p^7(\lambda sr)^k p^{k-1} \qquad [2]$$

Second alternative (**aB**$_2$): Alternative **aB**$_2$ follows the same strategy of forming three backcross lines as **aB**$_1$, but differs in the crossing steps. After backcross generations, selected males of one of the lines are crossed simultaneously with selected females of other two lines. Offspring heterozygous at two QTL regions are selected within both lines, and the individuals are mated to produce the final generation. From the final generation, individuals carrying the desired alleles at the three QTL regions are selected to start the intercross phase. Some of these selected individuals will be homozygous at the QTL region carried by the line used to cross with the other two lines. After the different crosses between lines, the expected number (N_2) of animals carrying the three QTL (including those homozygous at one QTL) is:

$$N_2 = 4n\lambda^3(1-s)^2 p^8(\lambda sr)^k p^{k-1}(p+2) \qquad [3]$$

Third alternative (**aB$_3$**): In alternative **aB$_3$** two lines are formed during the backcross, each carrying two QTL regions, that is, one QTL region is selected in both lines that are kept separate throughout the backcross phase. After the backcross generations, individuals heterozygous at two QTL from the two lines are crossed. From resulting offspring, animals carrying the three QTL regions are selected to start the intercrossing phase. Some of selected animals are homozygous at the QTL region common to the two lines.

From generation 1 up to generation k-1, the expected number (M_t) of animals heterozygous at two QTL regions for generation t within each line is $M_t = 2n\lambda sp^3(\lambda sr)^t p^{2(t-1)}$. At generation k, the expected number (N') of animals (males and females) heterozygous at two QTL for each line is $N' = 2n\lambda p^3(\lambda sr)^k p^{2(k-1)}$. After crossing the two lines, the expected number (N_3) of animals carrying the three QTL regions (including individuals homozygous at one) is:

$$N_3 = 4n\lambda^2(1-s)p^6(\lambda sr)^k p^{2(k-1)}(2+p) \qquad [4]$$

Fourth alternative (**aB$_4$**). The fourth alternative is based on two backcross lines: one carrying two QTL regions and the other carrying the third QTL region. At the end of backcross generations, selected males from the double carrier line are crossed with females selected from the single carrier line. From resulting offspring, individuals heterozygous at all three QTL regions are selected to start the intercrossing phase.

From generation 1 up to generation k - 1, the expected number (M_t') of animals heterozygous at two QTL in generation t is $M_t' = 2n\lambda sp^3(\lambda sr)^t p^{2(t-1)}$ for the double carrier line. For the single carrier line, the expected number (M_t'') of animals heterozygous at one QTL in generation t is $M_t'' = 2n\lambda sp^3(\lambda sr)^t p^{t-1}$. At generation k, the numbers are $N' = 2n\lambda p^3(\lambda sr)^k p^{2(k-1)}$ for the double carrier line and $N'' = 2n\lambda p^3(\lambda sr)^k p^{k-1}$ for the single carrier line. Crossing the two lines as described in previous paragraph results in the following number of individuals heterozygous at the three QTL regions:

$$N_4 = 2n\lambda^2(1-s)p^6(\lambda sr)^k p^{k-1} \qquad [5]$$

2. Intercross phase

2.1. Traditional intercross (**tI**)

Traditional intercross consists in mating individuals heterozygous at all three QTL and selected at the end of the backcross phase. From N heterozygous individuals at three QTL, $N(1 - s)$ are females which produce $N(1 - s)\lambda$ offspring in the first intercross generation. The expected total number (T_0) of homozygous animals at all three QTL regions is:

$$T_0 = N(1-s)\lambda p^6 \qquad [6]$$

2.2. Alternative intercross strategies

Two alternative intercross strategies (**aI**) are described and aim at maximizing the number of animals homozygous at all three QTL regions.

First alternative (**aI**$_1$). Once homozygous animals for all three QTL regions are selected at the first intercross, one can increase their number through a multiplication phase (Visscher and Haley 1999; Van der Waaij and Van Arendonk 2000). Alternative **aI**$_1$ consists in mating these homozygous animals, which results in an expected total number

$$T_1 = N(1-s)^2\lambda^2 p^6 \qquad [7]$$

of homozygous offspring, which is $(1 - s)\lambda$-fold the value in Eq [6]. There is no need for additional genotyping of offspring resulting from these matings, because only homozygous parents are used to produce the second intercross generations.

Second alternative (**aI**$_2$). When introgressing three QTL regions, some of the animals generated at the first intercross are carrying the three QTL regions. We can classify these animals in different groups as G1: animals homozygous at all three QTL; G2: animals homozygous at any two QTL and heterozygous at the third; G3: animals homozygous at any one of the three QTL and heterozygous at the other two; G4: animals heterozygous at the three QTL. There are more types of animals at the first intercross generation, but we restrict ourselves to these four groups because a cross within or between these groups can generate animals homozygous at three QTL regions. In the traditional intercross, only G1 animals are selected. However, animals in all four groups may be selected and used for a second intercrossing generation to increase the number of homozygous individuals at the

three QTL regions. From different mating combinations among these four groups of animals, the combination yielding the largest number of individuals homozygous at all three QTL regions is mating G1 males with females from the four groups (results not shown). The expected number of homozygous individuals at three QTL regions crossing animals from the four groups can be derived as follows.

After the first generation of intercrossing, the expected number of animals in G1 is $N(1-s)\lambda p^6 = U$ (see also Eq [6]). The expected number of animals in G2 is $6N(1-s)\lambda p^6 = 6U$. For G3, the expected number is $12N(1-s)\lambda p^6 = 12U$, and for G4 it is $8N(1-s)\lambda p^6 = 8U$. The expected number of animals homozygous at three QTL regions from different matings is $U(1-s)\lambda$, $6U(1-s)\lambda p$, $12U(1-s)\lambda p^2$, $8U(1-s)\lambda p^3$ for G1 males mated with G1, G2, G3 and G4 females, respectively. These matings result in an expected total number (T_2) of animals homozygous for all three regions equal to $U(1-s)\lambda(1+6p+12p^2+8p^3) = U(1-s)\lambda(1+2p)^3$. Replacing U by Eq [6] gives:

$$T_2 = N(1-s)^2\lambda^2 p^6(1+2p)^3. \qquad [8]$$

Comparing Eq [8] for $\mathbf{aI}_2$ with equation for **tI** (Eq [6]) and $\mathbf{aI}_1$ (Eq [7]) shows that the expected number of homozygous animals in $\mathbf{aI}_2$ is $(1+2p)^3\lambda(1-s)$ and $(1+2p)^3$ times that of **tI** and $\mathbf{aI}_1$, respectively.

3. Combining backcross and intercross phases (introgression strategies (IS))

Assuming that s varies around 0.5 and $\lambda \geq 3$, the second alternative of intercross strategy ($\mathbf{aI}_2$) yields more animals homozygous at all three QTL regions than **tI** and $\mathbf{aI}_1$. Therefore, only combinations involving $\mathbf{aI}_2$ and different backcross strategies will be studied here.

Combination of **tB** *and* $\mathbf{aI}_2$ ($\mathbf{IS}_0$). The expected total number (T_{IS_0}) of animals homozygous at all three QTL regions can be expressed as a function on n by replacing N in Eq [8] by N_0 in Eq [1]: $T_{IS_0} = 2n\lambda^3(\lambda s p^3 r)^k p^6(1-s)^2(1+2p)^3$.

For any number (z) of unlinked QTL regions to be introgressed, and using the introgression strategy $\mathbf{IS}_0$, T_{IS_0} can be expressed in a general way as:

$$T_{IS_0} = 2n\lambda^3\left(\lambda sp^z r\right)^k (1-s)^2 p^{2z}(1+2p)^z$$

Combination of $\mathbf{aB_1}$ *and* $\mathbf{aI_2}$ ($\mathbf{IS_1}$). Combining $\mathbf{aB_1}$ and $\mathbf{aI_2}$ results in an expected total number (T_{IS_1}) of animals homozygous at all three QTL given by Eq [8] where N takes the value of N_1 in Eq [2]: $T_{IS_1} = 4n\lambda^5\left(\lambda sr\right)^k p^{k-1} p^{13}(1-s)^4(1+2p)^3$.

Combination of $\mathbf{aB_2}$ *and* $\mathbf{aI_2}$ ($\mathbf{IS_2}$). In Eq [3], when developing $\mathbf{aB_2}$, two sub-populations were identified: individuals heterozygous at all three QTL, and individuals homozygous at one QTL and heterozygous at the other two QTL. Males from the second group are mated with females from both groups in the first intercross of $\mathbf{aI_2}$. The expected total number (T_{IS_2}) of animals homozygous at three QTL is given by:

$$T_{IS_2} = 4n\lambda^5\left(\lambda sr\right)^k p^{k-1} p^{13}(1-s)^4(1+2p)^2(3+2p).$$

Combination of $\mathbf{aB_3}$ *and* $\mathbf{aI_2}$ ($\mathbf{IS_3}$). Following the same strategy as in previous combination ($\mathbf{IS_2}$), the expected total number (T_{IS_3}) of animals homozygous at three QTL regions resulting from $\mathbf{aI_2}$ in combination with $\mathbf{aB_3}$ is:

$$T_{IS_3} = 4n\lambda^4\left(\lambda sr\right)^k p^{2(k-1)} p^{11}(1-s)^3(1+2p)^2(3+2p).$$

Combination of $\mathbf{aB_4}$ *and* $\mathbf{aI_2}$ ($\mathbf{IS_4}$). The combination of $\mathbf{aB_4}$ with $\mathbf{aI_2}$ results in an expected total number (T_{IS_4}) of animals homozygous at all three QTL given by Eq [8], where N takes the value of N_4 in Eq [5]: $T_{IS_4} = 2n\lambda^4\left(\lambda sr\right)^k p^{k-1} p^{12}(1-s)^3(1+2p)^3$.

4. Genotyping load evaluation

So far we have ignored the computation of the expected number of animals to be genotyped throughout the introgression process. This number (G_t) can be calculated from the expected number of animals with the desired genotype and the proportion of individuals with the desired genotype in each generation. The summation of these numbers over all generations of the introgression strategy gives the expected total number (G) of animals to be genotyped:

$$G = \sum_{t=1}^{k+\alpha} G_t$$

where $k + \alpha$ is the total number of generations in the introgression strategy.

The expected number of genotypings (W) depends on the expected total number of animals to be genotyped (G), the number of QTL region (z), and the number of markers (m) used to trace each QTL region. In the present evaluation, m is kept constant for each QTL region for simplicity. For the introgression strategy IS_0, the expected number of genotypings is simply $W = G \times z \times m$ because z is constant through the introgression process. For other introgression strategies, z must be considered at each generation within each line.

5. Cost analysis

Two major factors contributing to the cost (C) of an introgression strategy are considered: rearing cost and genotyping cost. Costs are derived for an experiment in mouse, but can be generalized to other species.

The rearing cost (R) is the cost of keeping all animals involved in the experiment for the time they are in use. These animals consist of all offspring produced and all animals that are selected as parents for the next generation. Females generated from F_1 up to generation k - 1 are not genotyped and not used in the backcross program, but in the case of mice they are kept until weaning. Assume c_1 is the cost of maintaining one animal for one generation and c_2 the cost of keeping an animal (including females) until the end of the selection process, then

$$R = c_1(G_s + B + F) + c_2 \times G_u$$

where G_s is the number of animals genotyped and selected through the introgression strategy, G_u the number of animals genotyped but not selected during the introgression with $G = G_s + G_u$, B the number of recipient animals kept for the purpose of the introgression program, and F the number of founder animals and F_1 males.

The genotyping costs (Y) include the costs related to animal sampling, DNA extraction and quantification, the cost of genotyping (PCR and analysis of results). Part of these costs is proportional to the number of genotypings (W). The other part increases by step (Darvasi and Soller 1994), that is, the cost is constant up to a given number of samples as a result of the genotyping equipment. In the present evaluation, we assume that 94 samples and two controls are analyzed simultaneously. Let c_3 be the variable cost of each genotyping, c_4 the cost of sampling and DNA preparation, and c_5 the cost of a set of one through 94 samples, then:

$$Y = c_3 \times W + c_4 \times G + c_5 \times m \sum_{t=1}^{k+\alpha} \sum_{i=1}^{L_t} z_{ti} [G_{ti} / 94]$$

where $[\cdot]$ is a ceiling function, and L_t is number of lines within generation t. Costs of introgression program are calculated as $C = R + Y$.

Results

The different backcross strategies described are all targeting 1000 individuals heterozygous at three QTL regions at the start of the intercross phase. While **tB** requires the highest number of genotypings, $\mathbf{aB_4}$ requires only half the number of genotypings of **tB** (Table 4.1). From Table 4.1, it can be concluded that $\mathbf{aB_4}$ results in the smallest number of animals to be genotyped. Numbers of genotypings in the four alternative backcross strategies are very similar. However, as a result of the number of lines within each strategy, the number of animals to be genotyped varies largely between alternatives. It can be noted that $\mathbf{aB_3}$ and $\mathbf{aB_4}$ have one generation less than other alternative backcross strategies.

Table 4.1 Effect of backcross strategies on number of animals to be genotyped and number of genotypings given a target number (N = 1000) of individuals heterozygous at all three QTL regions to start the intercross phase ($\lambda = 5$, $s = 0.6$, $k = 5$, $d = 20$cM, $r = 5$, $z = 3$)

Backcross strategies[a]	n[b]	No. of animals to be genotyped	No. of genotypings
tB	86.67	47,164	424,472
$\mathbf{aB_1}$	0.304	54,868	204,035
$\mathbf{aB_2}$	0.308	43,406	220,841
$\mathbf{aB_3}$	3.68	31,138	205,970
$\mathbf{aB_4}$	0.498	24,999	201,267

[a] **tB** = traditional backcross; **aB** = alternative backcross with subscripts indicating the type of alternative.
[b] n = half number of founder animals for each strain.

The expected numbers of animals homozygous resulting from different intercross strategies are given in Table 4.2. Starting with 1000 individuals, the traditional intercross (**tI**) results in nine animals homozygous at three QTL regions. Using one generation of multiplication ($\mathbf{aI_1}$) increases the expected number of homozygous animals by a factor $(1 - s)\lambda$, which is $0.4 * 5 = 2$. This alternative does not lead to an increase in the number of genotypings. The largest improvement, however, is made by the use of $\mathbf{aI_2}$ in which females carrying all three QTL regions are also used for the production of the second generation. In that case, the number of homozygous animals is 12 times larger than with **tI.** Applying $\mathbf{aI_2}$ requires some additional genotypings (Table 4.2).

Table 4.2 Effect of intercross strategies on the expected number of homozygous animals at three QTL regions, the number of animals to be genotyped and number of genotypings, starting with N = 1000 heterozygous individuals at all three QTL ($\lambda = 5$, s = 0.6, d = 20 cM)

Intercross strategies[a]	Expected no. of homozygous animals	Expected no. of animals to be genotyped	Expected no. of genotypings
tI	9	2,000	18,000
aI$_1$	18	2,000	18,000
aI$_2$	113	2,489	21,050

[a] **tI** = traditional intercross strategy; **aI$_1$** = first alternative intercross strategy; **aI$_2$** = second alternative intercross strategy.

The total number of animals that need to be genotyped to produce 100 animals homozygous at three QTL regions for different intercrossing strategies, combined with the traditional backcross strategy, is given in Table 4.3. The application of **aI$_2$** results in a large reduction in the expected number of animals to be genotyped and the number of genotypings. In the traditional introgression strategy (**tB + tI**), the number of animals to be genotyped is 522,332, of which 95.9% occur during the backcrossing stage (Table 4.3). With **aI$_2$**, the number of animals is reduced by 91.6% (Table 4.3). The differences between intercrossing alternatives in number of animals to be genotyped are also reflected in the number (n) of animals at the start of the introgression program and the total number of genotypings.

Table 4.3 Effect of traditional backcross strategy combined with alternative intercross strategies in an introgression program targeting 100 animals homozygous at three QTL ($\lambda = 5$, s = 0.6, d = 20 cM, r = 5, k = 5 and m = 3; in brackets are the expected proportions in the backcross phase).

Introgression strategies[a]	n[b]	No. of animals to be genotyped	Number of genotypings
tB + tI	920	522,332 (0.959)	4,700,984 (0.959)
tB + aI$_1$	460	261,166 (0.959)	2,350,492 (0.959)
tB +aI$_2$	76	43,844 (0.949)	393,402 (0.952)

[a] **tB** = traditional backcross strategy; **tI** = traditional intercross strategy; **aI$_1$** = first alternative intercross strategy; **aI$_2$** = second alternative intercross strategy.
[b] n = half number of founder animals for each strain.

Table 4.4 shows that introgression strategy **IS$_1$** has the largest number of animals to be genotyped, whereas it has a smaller number of genotypings and a lower cost than **IS$_0$**. The number of animals to be genotyped depends on r (number of females mated to one male) and λ (number of offspring per female). The introgression strategy, with the smallest number of animals to be genotyped, for a wide range of values for r and λ is shown in Fig. 4.1. **IS$_3$** appears to be the best alternative for $r \geq 3$ and $\lambda \geq 4$. However, when $\lambda \geq 100$ there is no difference between **IS$_2$** and **IS$_3$** (Fig. 4.1).

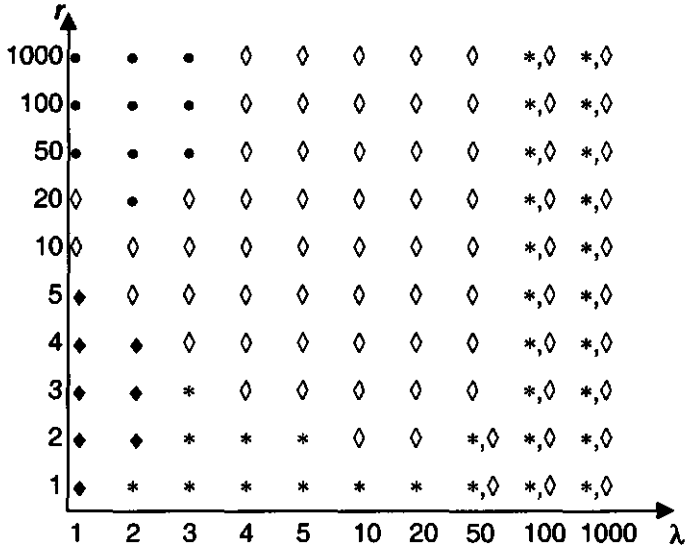

Fig. 4.1 Best introgression strategies (IS) distribution in terms of least number of animals to be genotyped given the number of offspring per female and the number of females per male when introgressing three QTL (♣ IS_0; • IS_1; * IS_2; ◊ IS_3; ♦ IS_4). Target number of homozygous animals at the three QTL is set to 100; $s = 0.6$, $d = 20$ cM, $k = 5$.

Table 4.4 Effect of introgression strategies on the total no. of animals to be genotyped, the no. of genotypings and the costs, targeting 100 homozygous animals at three QTL ($z = 3$, $\lambda = 5$, $s = 0.6$, $d = 20$ cM, $r = 5$, $k = 5$ and $m = 3$; in brackets are the expected proportions in the backcross phase)

Options[a]	α[b]	Total no. of animals to be genotyped	Number of genotypings	Estimated cost of introgression[c] (US$)
IS_0	2	43,844 (0.949)	393,402 (0.952)	464,433
IS_1	4	50,647 (0.956)	198,753 (0.906)	360,820
IS_2	4	19,167 (0.939)	100,499 (0.911)	146,872
IS_3	3	14,078 (0.917)	94,331 (0.905)	118,051
IS_4	3	24,273 (0.909)	196,309 (0.905)	268,430

[a] IS = introgression strategy where subscripts indicate the type of strategy.
[b] α is the number of generations after backcross steps; the observed differences between IS originate from corresponding alternative backcross strategies.
[c] Estimation based on an introgression in mice where c_1 = \$5.00; c_2 = \$1.00; c_3 = \$0.50; c_4 = \$2.50; c_5 = \$5.00; the equipment provision for depreciation and the labour cost have been ignored in these estimations.

Within the range given to r and λ, IS_3 remains the best introgression strategy in terms of the expected number of animals to be genotyped, the number of genotypings, and the cost of the introgression (Table 4.4). For parameters used in Table 4.4, IS_3 results in 20%

lower costs, 25% less animals to be genotyped and 6% less genotyping than the closest alternative, i.e. $\mathbf{IS}_2$. It is noteworthy that, compared to $\mathbf{IS}_0$, $\mathbf{IS}_3$ requires 76% fewer genotypings, requires 68% fewer animals to be genotyped, and costs 75% less.

Discussion

The main goal of designing alternative backcross and intercross strategies is to decrease the number of animals to be handled during the introgression of multiple QTL. As shown by the results in Table 4.1, the implementation of alternative backcross strategies results in about 50% reduction of the number of genotypings. In terms of number of animals to be genotyped, $\mathbf{aB}_1$ is worse than **tB,** and $\mathbf{aB}_4$ results in the smallest number of animals. On the basis solely of the backcross phase as an introgression program, $\mathbf{aB}_4$ is the best alternative to introgress three QTL. Results in Table 4.4 demonstrate, however, that it is also important to consider the intercross phase when planning an introgression program. The advantage of $\mathbf{aB}_3$ and $\mathbf{aB}_4$ in terms of number of required generations is economically important when considering agricultural species with long generation times.

Alternative backcross strategies applied in this paper are not restricted to the introgression of QTL coming from a single donor population. It is possible to use these alternative backcross strategies to introgress different QTL from two sources ($\mathbf{aB}_1$-$\mathbf{aB}_4$) or three sources ($\mathbf{aB}_1$ and $\mathbf{aB}_2$) to a single recipient breed.

Adding an extra intercross generation to the traditional intercross, that is, applying ($\mathbf{aI}_1$), results in an increase of the expected number of homozygous animals (Table 4.2). There will be an increase if $(1-s)\lambda > 1$, which implies that $\lambda \geq 3$, assuming s is in the range 0.4-0.6.

Making use of individuals carrying all three QTL regions and resulting from the first intercross ($\mathbf{aI}_2$) requires some additional genotyping (Table 4.2). This additional genotyping is overshadowed, however, by the relatively large decrease in the number of genotypings (Table 4.3). The increase in the expected number of animals homozygous at all three QTL regions (Table 4.2) is much larger than when only homozygous animals are used. The multiplier factor is $(1+2p)^3(1-s)\lambda$, as can be seen from comparing Eq [8] with Eq [6], and increases exponentially with z (z = 3 in this case). The use of types of individuals other than only homozygous animals generated at the first intercross, when extending to a second generation of intercrossing, is, therefore, an efficient and powerful means of implementing an introgression program, especially when the number of QTL regions is larger than one.

In this approach ($\mathbf{aI}_2$) of maximizing the use of certain types of animals from the first intercross, the number of G1 males might be a limiting factor mainly for species where the reproductive capacity is limited to natural mating. In a mice experiment, using males over a longer period of time can increase the number of matings. However, in species where artificial insemination (AI) is possible (for example, cattle, sheep, horse) or mating capacity of males is high (for example, poultry), the mating capacity of G1 males will not be a limiting factor. Restricting the number of males will have a negative effect on the level of inbreeding in subsequent generations.

At low values of r or λ, none of the introgression strategies remains constantly the best approach (Fig. 4.1). The introgression strategy $\mathbf{IS}_3$ appears, however, to be the best introgression alternative, requiring fewer animals to be genotyped when $r \geq 3$ and $3 < \lambda \leq 50$. When $\lambda \geq 50$, the difference between $\mathbf{IS}_2$ and $\mathbf{IS}_3$ becomes small and disappears from $\lambda = 100$ for $r \geq 3$. The advantage of $\mathbf{IS}_2$ and $\mathbf{IS}_3$ results mainly from the fact that the proportion of homozygous animals for all three QTL regions in the first intercross generation is p^{-1} times higher than in the other introgression strategies. These differences result from some animals which, carrying the three QTL regions and undergoing the intercross, are already homozygous at one region. This reduces the expected number of animals to be genotyped at this level to meet the target number of homozygous animals at three QTL regions, and consequently reduces also the number of animals in previous generations.

Whereas it is relatively easy to increase the number of females (r) per male by means of artificial insemination (AI) in some species, the increase of r will be limited in species where AI is not usually used. Moreover, the number of offspring per female cannot be increased indefinitely, even by means of multiple ovulation and embryo transfer (MOET) techniques.

In order to decrease the minimum number of individuals to be genotyped through an introgression program, Hospital and Charcosset (1997) have described a so-called pyramidal design, which is similar to alternative $\mathbf{aB}_1$. As indicated by these authors, the pyramidal design and $\mathbf{aB}_1$ trace the QTL individually and make possible a higher background selection intensity in the backcrossing phase on animals already selected for the presence of the QTL. However, according to current results and in absence of a background selection, this option is not optimal within the restriction $r \geq 3$ and $\lambda \geq 4$. Including four unlinked QTL regions in our model (results not shown) illustrates that a backcross phase with two double QTL carrier lines is more advantageous than a backcross phase with four lines of single QTL carriers.

The number of genotypings is a product of the number of animals, the number of loci involved, and the number of markers per locus. Once the number of animals is minimized by the design, the number of genotypings can be reduced by a sequential genotyping at each step of the introgression process whenever it is possible. Genotyping and selection are performed at one locus, then for a second locus, and so on. Applying this approach, for example, to $\mathbf{IS}_0$ (three QTL regions, d = 20 cM), the expected number of genotypings is reduced by 47% $\left[1-\left(1+p+p^2\right)/3\right]$ at each step of the backcross phase, and by 60% $\left[1-\left(1+p^2+p^4\right)/3\right]$ at the first intercross generation. These reductions can even be improved when the sequential genotyping is extended to the markers at each QTL region, under the condition that the number of markers per region is at least two. Because markers at one locus are dependent, it is more likely that animals carrying one of them will also carry others. A further reduction in genotyping is also achievable during crosses between lines within the backcross phase and through the intercross phase by applying a mating ratio between males and females. This means that fewer males will be required to be genotyped.

Costs are important factors in decision making. The cost derivation method described in this paper considers the stepwise nature of increasing costs but does not include the cost c_3 of sampling and DNA extraction/storage in the rearing cost (R) [as done by Darvasi and Soller (1994)], but rather in the genotyping cost. Integrating such an introgression program in a commercial enterprise could remove the rearing cost, leaving only the genotyping cost (Y) as the introgression cost (Visscher and Haley 1999). Results in Table 4.4 show how much can be saved when $\mathbf{IS}_3$ is used to introgress three regions. Switching from $\mathbf{IS}_0$ to $\mathbf{IS}_3$ results in 75% cost saving. The savings would be higher if the QTL positions were accurately known, but in most studies these positions are known in a larger confidence interval. In these conditions, it is advisable to confirm the presence of target QTL (Spelman and Bovenhuis 1998) before starting an MAI experiment to avoid a loss of resources.

In each of the introgression strategies described in the present study, the adaptive breeding strategy developed by Koudandé and coworkers (1999) may also be applied. When the target number of animals predicted in a backcross generation is not obtained, one can increase r in the current backcross or decide to genotype and select females in the following generations. A marker-assisted introgression experiment is under way at the International Livestock Research Institute (ILRI), Nairobi, and $\mathbf{IS}_3$ is the introgression strategy in use. In this program, the adaptive breeding strategy is applied to restore numbers to the model predictions.

For experimental purposes, the introgression strategies developed here to introgress three QTL regions provide opportunities that can be exploited to study different aspects of the expression of QTL in the recipient background. Within the population generated at the first generation of intercross, there are individuals already genotyped which are homozygous for donor QTL region at one or two of the three QTL, and one needs only to select them. Examining the phenotypes of these individuals allows researchers to assess the effect of each QTL, the effect of combinations of the QTL two by two, and to detect any interactions (Frankel and Schork 1996). The effect of the background genotype on the expression of the three QTL can also be evaluated by comparing the introgressed animals with the three QTL to the donor line. Conversely, the recipient line can be compared to the introgressed animals carrying the recipient alleles at the three QTL.

Acknowledgements

The authors are grateful to Dr. Henk Bovenhuis and Piter Bijma for their helpful comments and suggestions. They would like to acknowledge The Netherlands Foundation for the Advancement of Tropical Research (WOTRO) and the International Livestock Research Institute (ILRI) for financial support. O.D.K. acknowledges the leave provided by INRAB (Institut National des Recherches Agricoles du Bénin) to carry out this research. This is ILRI Publication No. 990133.

References

Darvasi A, Soller M (1994) Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. *Theor Appl Genet* 89, 351-357

Frankel WN, Schork NJ (1996) Who's afraid of epistasis ? *Nat Genet* 14, 371-373

Gama LT, Smith C, Gibson JP (1992) Transgene effects, introgression strategies and testing schemes in pigs. *Anim Prod* 54, 427-440

Groen AF, Smith C (1995) A stochastic simulation study of the efficiency of marker-assisted introgression in livestock. *J Anim Breed Genet* 112, 161-170

Hospital F, Charcosset A (1997) Marker-assisted introgression of quantitative trait loci. *Genetics* 147, 1469-1485

Kashi Y, Hallerman E, Soller M (1990) Marker assisted selection of candidate bulls for progeny testing programmes. *Anim Prod* 51, 63-74

Kemp SJ, Iraqi F, Darvasi A, Soller M, Teale AJ (1997) Localization of genes controlling resistance to trypanosomiasis in mice. *Nat Genet* 16, 194-196

Koudandé OD, Thomson PC, Van Arendonk JAM (1999) A model for population growth of laboratory animals subjected to marker assisted introgression: how many animals do we need? *Heredity* 82, 16-24

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124, 743-756

Markel P, Shu P, Ebeling C, Carlson GA, Nagle DL, Smutko JS, Moore KJ (1997) Theoretical and empirical issues for marker-assisted breeding of congenic mouse strains. *Nat Genet 17*, 280-283

Spelman RJ, Bovenhuis H (1998) Moving from QTL experimental results to the utilisation of QTL in breeding programmes. *Anim Genet* 29, 77-84

Van der Waaij EH, Van Arendonk JAM (2000) Introgression of genes responsible for disease resistance in an otherwise for production selected cattle population: genetic and economic consequences. *Anim Sci* 70, 207-220

Visscher PM, Haley CS (1999) On the efficiency of marker-assisted introgression. *Anim Sci* 68, 59-68

Visscher PM, Haley CS, Thompson R (1996) Marker-assisted introgression in backcross breeding programs. *Genetics* 144, 1923-1932

Wakeland E, Morel L, Achey K, Yui M, Longmate J (1997) Speed congenics: a classic technique in the fast lane (relatively speaking). *Immnol Today* 18, 472-477

Yancovich A, Levin I, Cahanar A, Hillel J (1996) Introgression of the avian naked neck gene assisted by DNA fingerprints. *Anim Genet* 27, 149-155

CHAPTER V

Trypanotolerance QTL introgression in mice: experimental results

Olorounto D. KOUDANDÉ[1], Johan A. M. van ARENDONK[1], Henk BOVENHUIS[1], John P. GIBSON[2], and Fuad IRAQI[2]

[1] *Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands*

[2] *International Livestock Research Institute (ILRI), P.O. Box 30709 Nairobi, Kenya*

Abstract

A marker-assisted introgression (MAI) experiment was conducted to transfer trypanotolerance QTL from a donor mouse strain, C57BL/6, into a recipient mouse strain, A/J. The objective was to assess the effectiveness of such introgression using genetic markers and evaluate the effect of each of the three QTL within the new background genotype. We used a backcross strategy that consisted in selecting two lines, each carrying two of the donor QTL alleles through the backcross (BC) phase. At the fourth BC generation, single carrier animals were selected to proceed with the intercross phase. Given the genotype on chromosomes 1, 5 and 17, ten groups of mice plus two controls (A/J and C57BL/6) were constituted at the end of the intercross phase and underwent the challenge with Trypanosoma congolense. Sequential genotyping was applied to BC_3 animals. In this paper, we report on the results obtained during the first 90 days post-inoculation. The results showed that markers were transmitted from a generation to another and could be traced through the breeding phase. During this phase, some bottlenecks were encountered: failure of selected animals to survive or to reproduce, availability of few animals with the desired genotype, and availability of animals from only one sex, important during the intercross phase. Sequential genotyping reduced the genotyping load by more than 50%. Females had longer mean survival time than males ($P < 0.001$). Significant effects were found for all three chromosomes. The additive effect of QTL measured in days of survival was 10.0 on chromosome 1, 7.0 on chromosome 5, and 6.0 on chromosome 17. The proportion of the donor background with its regression coefficient of 58.56 ($P=0.0009$) had a remarkable effect on survival time. The donor's QTL alleles tended to act recessively on chromosomes 1 and 17 whereas alleles acted additively on chromosome 5. The present experiment shows that introgression is feasible and effective in relatively short time with laboratory animals.

Introduction

Trypanosomosis is beyond doubt the most important constraint to livestock development in the sub-humid and non-forested portions of the humid zone of Africa. The disease costs approximately US $1340 million per year for livestock producers in Africa (Kristjanson et al. 1999). This cost excludes losses due to reduction of manure availability and the inability to use draught power.

N'Dama and West African Shorthorn cattle (Baoulé, Muturu, Lagune) are recognized for their ability to withstand the effect of trypanosomes infection and to remain productive in areas where trypanosomosis prevents the presence of other cattle types, or significantly reduces their productivity (Murray and Trail 1984; Trail et al. 1989). This ability to

withstand trypanosomes infection, called trypanotolerance, is an innate feature of the long horn N'Dama and other Shorthorn cattle from West Africa (Roberts and Gray 1973; Roelants 1986; Doko et al. 1991).

At the International Livestock Research Institute (ILRI) in Nairobi-Kenya, a linkage study is underway on an F_2 generation of a cross between N'Dama and Boran cattle to identify genes or QTL involved in trypanotolerance (Teale 1993). The detection and the identification of genes related to trypanotolerance as well as some linked markers will open the way for introgression of these genes in more susceptible cattle.

Various laboratory inbred mouse strains show variation in resistance to *Trypanosoma congolense* infection. Among these laboratory mouse strains, the C57BL/6 strain appears to be one of the most resistant with mean survival time of 110.2 days whereas the A/J strain, with mean survival of 15.8 days, appears to be the least resistance (Morrison et al. 1978). Although the survival times vary considerably between strains following challenge, the infection with *T. congolense* usually results in death of all strains of mice denoting the relative character of resistance to trypanosomosis (Teale et al. 1999).

Based on two F_2 populations, (A/J x C57BL/6) and (Balb/c x C57BL/6), Kemp et al. (1996, 1997) showed that three chromosomal regions were associated with trypanotolerance in mice. Recently Iraqi et al. (2000) succeeded to fine-map QTL influencing survival time on chromosome 17 in mice following challenge with *T. congolense* to a small confidence interval (0.9 cM) using an F_6 advanced intercross line following Darvasi and Soller (1995). From the same study, the confidence interval for the QTL on chromosome 5 was 12 cM. This study, however, resolved the previous QTL on chromosome 1 (Kemp et al. 1997) into three distinct trypanoresistance QTL.

The marker density now available on the mouse genome and the identification of markers linked to these QTL allow a marker-assisted introgression experiment in mice. For such an experiment, mice of course are more convenient in terms of costs and generation-interval than cattle, and can serve as model.

Marker-assisted introgression is a crossbreeding program that aims at incorporating genes from a donor to a recipient animal through a backcross design. Marker-assisted introgression consists of creation of F_1 generation from founder animals, followed by a number of backcross generations and completed with an intercross phase to fix the introgressed genes (Soller and Plotkin-Hazan 1977; Groen and Smith 1995).

The present paper aims at investigating the effectiveness of a marker-assisted introgression experiment of trypanotolerance QTL alleles from the mouse strain C57BL/6 (donor) into the mouse strain A/J (recipient). The experiment consists of two parts, first the breeding phase which is the effective phase of the introgression, and second the challenge phase designed to verify whether the QTL monitored by means of markers are really

present in the synthetic mice. From the challenge phase, we evaluate the significance and magnitude of each QTL effect in the recipient background genotype, the existence of interactions and the effect of the residual donor background genotype in recipient animals.

Materials and methods

Breeding program

Inbred mouse strains A/J (OlaHsdnd) and C57BL/6 (OlaHsd) purchased from Harlan UK Ltd., Bicester, U.K., were maintained as pure strains at ILRI for research purposes. In current experiment, C57BL/6 being resistant was the donor, and the sensitive A/J was the recipient. For the breeding program, a reciprocal cross was performed between founder parents (10 males C57BL/6 x 10 females A/J and 10 males A/J x 10 females C57BL/6) to produce F_1 animals. Next, F_1 males were crossed to A/J females and A/J males were crossed to F_1 females to produce the first backcross generation (BC_1). Male mice from BC_1 were genotyped and two types of mice were selected, each carrying two of the three QTL regions, that is, mice carrying trypanotolerance alleles for the QTL regions on chromosomes 1 and 5, and mice carrying trypanotolerance alleles for the QTL regions on chromosomes 5 and 17. These two types of mice were maintained separately along the backcross phase and were designated heterozygous lines $L_{1,5}$ and $L_{5,17}$ as described by Koudandé et al. (2000). Selected BC_1 males were backcrossed to A/J females to produce BC_2 mice within each of the two lines. From BC_2 onwards, males and females were genotyped and selected as parents for the next generations.

In addition to lines $L_{1,5}$ and $L_{5,17}$, three other types of heterozygous "single carrier" mice were selected at BC_4, that is first, mice carrying donor alleles at the QTL region on chromosome 1 and the recipient alleles at the QTL region on chromosomes 5 and 17 (L_1). The second type (or L_5) was mice carrying donor alleles at the QTL region on chromosome 5 and the recipient alleles at the QTL region on chromosomes 1 and 17. The third type (or L_{17}) was mice carrying donor alleles at the QTL region on chromosome 17 and the recipient alleles at the QTL region on chromosomes 1 and 5. The selected lines $L_{1,5}$ and $L_{5,17}$ were used for the purpose of simultaneous introgression of the three QTL regions.

After BC_4, an intercrossing was performed within each of the new lines L_1, L_5 and L_{17}, and aimed at producing homozygous individuals for the donor alleles at the QTL region. Because of the limited number of selected females in L_{17}, males from L_{17} were crossed with females from $L_{5,17}$ to generate some mice homozygous for the donor allele at chromosome 17. A second generation of intercrossing was performed to increase the

Table 5.1 Different groups of mice to be challenged, their genotype profile (A=A/J allele (*recipient*), C=C57BL/6 allele (*donor*)) and their origin

Groups	Chr.1	Chr.5	Chr.17	Origin
1	CC	AA	AA	Homozygous on chr.1 for donor alleles (from L_1)
2	AC	AA	AA	Cross of group 1 with the recipient A/J
3	CC	AC	AC	Cross of group 1 with the donor C57BL/6
4	AA	CC	AA	Homozygous on chr.5 for donor alleles (from L_5)
5	AA	AC	AA	Cross of group 4 with the recipient A/J
6	AC	CC	AC	Cross of group 4 with the donor C57BL/6
7	AA	AA	CC	Homozygous on chr.17 for donor alleles (from L_{17})
8	AA	AA	AC	Cross of group 7 with the recipient A/J
9	AC	AC	CC	Cross of group 7 with the donor C57BL/6
10	AA	AA	AA	Internal control selected from intercrossed L_1, L_5, L_{17}
11	AA	AA	AA	Control recipient line A/J from stock
12	CC	CC	CC	Control donor line C57BL/6 from stock

number of homozygous mice in each line by primarily crossing selected homozygous mice either with each other or with heterozygous mice of the same line, as well as crossing heterozygous mice. Meanwhile, mice that were homozygous for the recipient alleles at all three QTL regions were selected as internal controls (Group 10 in Table 5.1). In the last generation, 12 groups of mice were set up for the challenge phase (Table 5.1). Nine of the 12 groups were derived from the homozygous mice obtained in lines L_1, L_5 and L_{17}. All females from the three lines were crossed with few corresponding males to produce homozygous mice for groups 1, 4 and 5 in Table 5.1. The remaining males from these three lines were used for backcrossing to both parental lines to generate groups 2, 5 and 8 (backcross to A/J), and groups 3, 6 and 9 (backcross to C57BL/6).

Genotyping

Total genomic deoxyribonucleic acid (DNA) was extracted from tail tissue collected on three weeks old mice using the conventional method described by Sambrook et al. (1989). The extracted DNA was diluted, quantified and each DNA solution was adjusted to 0.05 mg/ml for polymerase chain reaction (PCR).

Three fluorescent-labeled primers for microsatellite marker amplification (Research Genetic Inc., Huntsville, AL, USA) were used per each QTL region to assess the alleles inherited by each mouse from BC_1 through BC_4. These microsatellite markers were D1Mit87, D1Mit217 and D1Mit60 for QTL region on chromosome 1, D5Mit200, D5Mit113 and D5Mit10 for QTL region on chromosome 5, D17Mit29, D17Mit16 and D17Mit11 for QTL region on chromosome 17. From the first intercross onwards three other markers, that is D5Mit58, D5Mit201 and D5Mit157, replaced D5Mit113 resulting in

a total of five markers investigated on QTL region on chromosome 5. The size of the QTL region was 5 cM on chromosome 1, 18 cM on chromosome 5 and 7 cM on chromosome 17 (Iraqi et al. 2000). All microsatellites were fixed at alternative alleles for the two mouse strains used to start the experiment.

The amplification of markers was performed according to the supplier recommendations using a thermocycler PTC100 (MJ Research, Inc., USA). PCR products were analyzed on a 4.25% polyacrylamide gel using an automated DNA sequencer ABI 377 (Perkin Elmer). Subsequently the tracking of the gel was checked and adjusted manually and analyses were performed using Genescan™ version 2.1 and Genotyper™ version 2.0 software (Applied Biosystems 1996).

At BC_3 all generated mice were sequentially genotyped as described by Koudandé et al. (2000) and Van der Waaij and Van Arendonk (2000), that is, selecting animals based on genotyping on a first marker, then on a second and so on. Genotyping was performed first on the target region of chromosome 5 because this region was common to both lines $L_{1,5}$ and $L_{5,17}$. Heterozygous animals for all three markers on this chromosome were selected and were typed for one marker at a time on chromosome 1 or 17 according to the line. Selection was done after each marker genotyping before proceeding with the next marker.

Challenge with T. congolense

Two irradiated rats were inoculated with *T. congolense*, clone IL1180 (Masake et al. 1983). Ten days post inoculation, the parasitaemia in rats was high enough to enable parasite collection from rat blood using a column of DEA cellulose. The infecting solution is adjusted at $0.5*10^5$ parasites per ml so as 10^4 parasites were injected to each mouse (intra-peritoneal) of the twelve groups under a volume of 0.2 ml of solution. The parasitaemia in mice was checked from the fourth through the fifteenth day post-inoculation by means of fresh blood smears observed through microscope. Only data from mice that tested positive for infection during this checking period were used for subsequent analyses. The numbers of days of survival i.e. the numbers of days post-inoculation before death of animals were recorded and analyzed. Two challenge operations were done at one-month interval. In this paper we report on results obtained during the first 90 days post-inoculation. All survival time for animals alive after 90 days was set equal to 91 days.

Statistical analysis

Survival analysis was performed using LIFETEST procedure in SAS (1990). This procedure computes the estimates of the survival function by the product-limit method also

called Kaplan-Meier method. It consists in calculating the survival probabilities for each group of mice from the numbers of mice alive on a daily base after challenge. The log-rank test was used to compare survival curves across different groups of mice (Kalbfleisch and Prentice 1980).

Analysis of QTL effects was performed with SAS procedure GLM using the following linear model:

$$Y_{ijklmnr} = \mu + Q1_i + Q5_j + Q17_k + S_l + \beta_1 X1_m + \beta_2 X2_n + e_{ijklmnr}$$

where $Y_{ijklmnr}$ was the survival time of individual r, Q1 corresponded to genotype effect of QTL on chromosome 1, Q5 was the genotype effect of QTL on chromosome 5, and Q17 was the genotype effect of QTL on chromosome 17. Genotype effects consisted of three levels: AA, AC, CC. S_l was the effect of sex, X1 was age at infection, X2 was the proportion of donor's background genotype, β_1 and β_2 are regression coefficients and $e_{ijklmnr}$ the residual error term. For the linear model, the control groups of pure lines (A/J and C57BL/6) were not included in this analysis.

Dominance effects were assessed by contrasting heterozygotes with half the sum of the estimates of the two homozygotes. To compare means we used Tukey-Kramer method, which adjusts for multiple comparisons.

Results

Breeding

The first 53 F_1 animals (30 males and 23 females) resulting from the reciprocal cross between the founder mouse strains were backcrossed to the recipient line A/J. The selection results based on the genotyping data were summarized in Table 5.2. Note that there were only three females within the 17 selected heterozygous mice in L_{17} at BC_4. The first intercross within lines L_1, L_5 and L_{17} followed by the genotyping resulted in 16 selected homozygous mice (9 males and 7 females) in L_1, two homozygous males in L_5 and four homozygous mice (2 males and 2 females) in L_{17}. The four homozygous mice in L_{17} resulted from the cross L_{17} x $L_{5,17}$ and not from the intercross of L_{17}. There were no homozygous females in L_5 and few homozygous mice were selected in L_{17}. To overcome this problem, heterozygous mice (males and females) were selected for a second generation of intercrossing. The second intercrossing generation followed by the genotyping resulted in 53 selected homozygous animals in L_1, 46 in L_5 and 26 in L_{17}.

From these homozygous mice, all females and part of the males were crossed to produce mice for groups 1, 4 and 7 as defined in the materials and methods section. The other part of males was backcrossed to females A/J to produce mice for groups 2, 5, 8 and to females C57BL/6 to produce mice for groups 3, 6, and 9. The respective numbers of challenged animals from group 1 through group 12 are 60, 54, 54, 51, 15, 62, 14, 27, 17, 59, 60 and 60. Offspring produced from L_{17} homozygous for the donor alleles did not survive after weaning. Challenged animals in group 7 resulted from the second intercross generation, i.e. from the cross between two heterozygous animals at the QTL region on chromosomes 17.

Table 5.2 Observed and expected frequencies of the desired genotype through the introgression phases (number of selected mice)

Steps[a]	Lines	No. of animals genotyped	Frequencies	
			Expected[b]	observed
BC_1	$L_{1,5}$	164	0.1986	0.0609 (10)
	$L_{5,17}$	164[c]	0.1947	0.0548 (9)
BC_2	$L_{1,5}$	68	0.1986	0.1912 (13)
	$L_{5,17}$	90	0.1947	0.1555 (14)
BC_3	$L_{1,5}$	104	0.1986	0.1250 (13)
	$L_{5,17}$	206	0.1947	0.0919 (16)
BC_4	L_1	299[d]	0.1853*	0.1037 (31)
	L_5	457[e]	0.1841*	0.1138 (52)
	L_{17}	158[f]	0.1819*	0.1076 (17)
	$L_{1,5}$	299	0.1853*	0.1170 (35)
	$L_{5,17}$	158	0.1819*	0.1582 (25)
IC_1	L_1	93	0.2262	0.1720 (16)
	L_5	134	0.1744	0.0149 (2)
	L_{17}	16	0.2173	0.0000 (0)
	L_{17} X $L_{5,17}$	25	0.0977	0.1600 (4)
IC_2	L_1	53	1.0000	1.0000 (53)
	L_5	368	0.1744	0.1250 (46)
	L_{17}	97	0.2173	0.2680 (26)

L = line of mice with subscripts indicating the chromosomes which carry the donor allele.
a: BC = backcross, IC = intercross; b: Derived after Koudandé et al. (1999 & 2000); c: these mice are the same as in $L_{1,5}$ in BC_1; d: same mice as $L_{1,5}$ in BC_4; e: sum of mice from $L_{1,5}$ and $L_{5,17}$ in BC_4; f: same mice as in $L_{5,17}$ in BC_4.
* See derivation in annex.

Sequential genotyping

Sequential genotyping were performed on mice from line $L_{1,5}$ according to the following order of markers: D5Mit113, D5Mit10, D1Mit60, D1Mit217, D1Mit87 and D5Mit200. Starting with 104 mice, 50, 47, 18, 16, 16 and 13 mice were respectively selected for the donor allele after each marker genotyping. The respective numbers of mice genotyped for each marker were 104, 50, 47, 18, 16 and 16 resulting in 60% saving compared to a conventional genotyping using all 104 mice against the six markers. In a first batch of mice in $L_{5,17}$, 80 mice were genotyped for D5Mit113, 30 for D5Mit10, 27 for D17Mit11, 13 for D17Mit29, 9 for D17Mit16 and 9 for D5Mit200. For a second batch, 94 mice were genotyped for D17Mit11, 36 for D5Mit10, 19 for D17Mit16, 19 for D5Mit200, 11 for D5Mit113 and 11 for D17Mit29. The savings were 65% and 66% respectively for the first and second batches in $L_{5,17}$. The number of selected mice after the last marker was six for the first batch and 10 for the second batch.

Challenge experiment

One mouse died three days after challenge even before infection control had started and, therefore, was eliminated from the analysis. At day 15 post-inoculation, 19 animals were still not parasitaemic and were discarded from following analyses. The infection rate was 96.4 per cent over 533 inoculated mice. The non-infected mice did not originate from a particular genotype group.

Table 5.3 Status and mean survival of different groups of mice following lifetest analysis (A=A/J allele, C=C57BL/6 allele)

Group	Genotype			Back-ground[1]	Infected mice (no.)	Censored mice (no.)	Mean survival (days)	s.e. mean survival
	Chr.1	Chr.5	Chr.17					
1	CC	AA	AA	0.031	60	1	57.917	3.308
2	AC	AA	AA	0.015	54	3	43.648	3.733
3	CC	AC	AC	0.515	52	41	87.192	0.642
4	AA	CC	AA	0.031	49	3	47.347	4.597
5	AA	AC	AA	0.015	13	2	47.308	5.970
6	AC	CC	AC	0.515	56	53	89.321	0.590
7	AA	AA	CC	0.031	14	1	45.286	5.733
8	AA	AA	AC	0.015	27	3	36.556	6.141
9	AC	AC	CC	0.515	13	12	90.000	-
10	AA	AA	AA	0.031	58	0	34.241	3.406
A/J	AA	AA	AA	0.000	60	0	28.650	3.006
C57BL/6	CC	CC	CC	1.000	58	6	69.259	1.629

[1] Estimated donor's background genotype according to Stam and Zeven (1981), and Young and Tanksley (1989)

The survival data of all groups is summarized in Table 5.3. The overall survival function across groups is plotted in Figure 5.1. In total 125 animals were censored (i.e. still alive at day 90) while 389 had died. At 90 days post-challenge, few animals had died in groups 3, 6 and 9. In Figure 5.1, except groups 3, 6 and 9, all survival curves laid between the controls i.e. the donor and the recipient groups of mice. Groups 3, 6 and 9 showed longer survival time than the donor line C57BL/6 and the log-rank test confirmed this result for group 3 ($\chi^2 = 5.89$; $P < 0.025$) and group 6 ($\chi^2 = 6.92$; $P < 0.01$). The survival curve of group 9 was further away from the donor curve but was not statistically different ($\chi^2 = 2.32$; $P > 0.10$). No significant differences in survival curve were observed among groups 1, 4 and 7, and between each of these groups and the donor line (group 12). The two control groups of inbred lines, however, showed significant difference between their survival curves ($\chi^2 = 4.86$; $P < 0.05$). There was also significant difference ($\chi^2 = 4.24$; $P < 0.05$) between the survival function of donor line and that of the internal control (group 10). No significant difference was found between the recipient line (group 11) and the internal control (group 10).

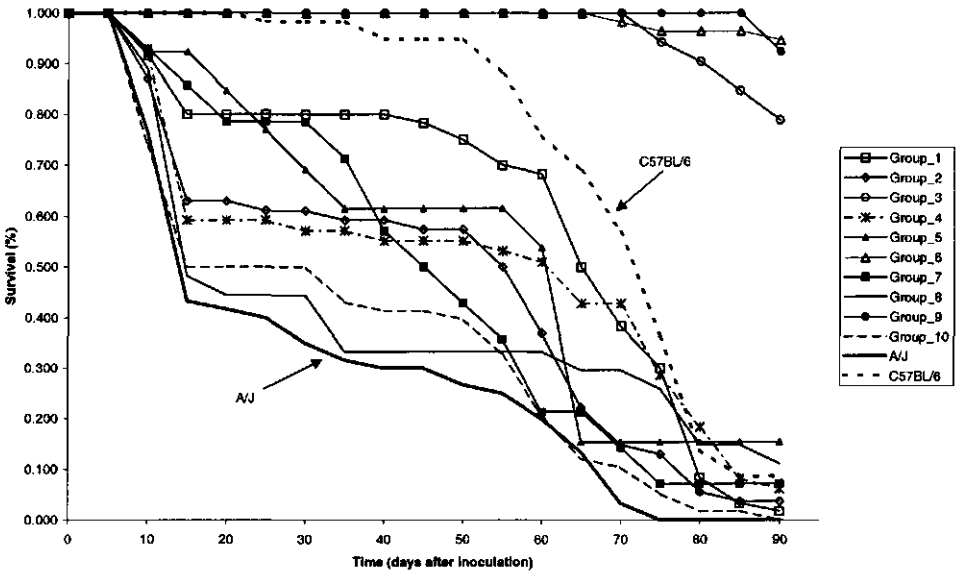

Figure 5.1 Survival of different genotype groups of mice following *Trypanosoma congolense* inoculation

Fitting the linear model, QTL on chromosome 17 (Q17) was not significant at 5% level (Table 5.4). Factors cage and challenge (two distinct challenge operations were conducted) were excluded from the model because not showing significant effect. QTL on chromosome 1 (Q1) accounts for 42% of the total sum of squares of the model whereas QTL on chromosome 5 (Q5) and Q17 account for 40% and 9% respectively. Except Q17,

all other factors included in the model were highly significant ($P \leq 0.005$). Among these factors, Q1 seemed the most significant with $P = 0.0001$. The survival estimates in Table 5.5 showed that Q1 had the largest effect whereas Q17 had the least effect. Q5 showed intermediate effect compared to Q1 and Q17. Females had longer mean survival time than males. The regression coefficient of the background genotype (58.56) showed that this factor had a big impact on survival time.

Table 5.4 Fitting the linear model

Source of variation	d.f.	Mean Square	F	P
Q1	2	7092.28	13.31	0.0001
Q5	2	2867.14	5.38	0.0050
Q17	2	1395.70	2.62	0.0742
Sex	1	6293.37	11.81	0.0007
Age	1	5368.96	10.08	0.0016
Background	1	5943.19	11.15	0.0009

Q1= QTL on chromosome 1; Q5= QTL on chromosome 5; Q17= QTL on chromosome 17.

Table 5.5 Survival estimates of different classes within factors

Factors	Class	Survival estimates	SE of difference
	AA	-19.827	4.159
Q1	AC	-14.157	3.937
	CC	0.000	-
	AA	-13.683	4.443
Q5	AC	-7.950	4.413
	CC	0.000	-
	AA	-11.939	5.987
Q17	AC	-11.448	5.426
	CC	0.000	-
Sex	Female	8.124	2.364
	Male	0.000	-
Age	-	-1.850*	0.583
Background	-	58.562*	17.535

Q1= QTL on chromosome 1; Q5= QTL on chromosome 5;
Q17= QTL on chromosome 17; * Regression coefficient

Based on least-square analysis, there was strong evidence ($P < 0.01$) that the mean survival time of homozygous mice for donor's QTL allele on chromosome 1 and 5 was

longer than the mean survival time of homozygous mice for recipient's QTL allele. This difference was not statistically significant on chromosome 17 (Table 5.4). For all three QTL, none of the mean survival times of heterozygotes was significantly different ($P > 0.42$) from those of homozygotes for the recipient QTL allele, though they are larger except on chromosome 17. The donor's QTL alleles showed recessive effect on chromosomes 1 and 17, however, this recessive effect was not statistically significant. The effect of the QTL allele was additive on chromosome 5 (Table 5.5). On chromosome 17, the estimated genotype effect of AA (-11.94 days) is almost identical to that of AC genotype (-11.45 days) consistent with complete recessive effect of the donor allele.

Discussion

Current experiment is unique because it addresses introgression of chromosomal regions (or QTL), and not introgression of an identified gene as in previous reports in animal species. Many examples of introgression experiment have been reported in plant breeding (e.g., De Vries et al. 1992; Oertel and Matzk 1999; Khrustaleva and Kik 2000; Lim et al. 2000). Yancovich et al. (1996) reported on an experiment aiming at introgressing the avian naked neck gene focusing on genomic selection to speed up the recipient genome recovery. In the same way, Markel et al. (1997) gave an experimental support to the rapid recovery of the recipient genome using the introgression of the *Apoe null* allele in different inbred mice. None of these experiments reported on the intercross phase, which is a marked difference with current paper. Another feature of our experiment is that QTL have been introgressed relying solely on genetic markers without any phenotypic measurements. This experiment gives a strong support to theories about the use of genetic markers, in particular the use of linked markers to trace QTL in a breeding program.

From the breeding program, offspring produced from homozygous individuals for donor's QTL allele on chromosome 17 fail to survive after weaning. Post-mortem investigation did not reveal any explanation. Observations on animals when still alive showed that animals grew slowly, became weaker and weaker, and finally died. The exact cause is not known but the results suggest that the QTL on chromosome 17, or some linked genes, may have a negative interaction with the new background genotype resulting in the depression of survival after weaning. Whatever the cause of the poor survival in group 7, if expressed during the challenge, it may have contributed to the rather low estimates of the effects of the chromosome 17 QTL on survival here when compared to earlier studies (Kemp et al. 1996, 1997; Iraqi et al. 2000).

Results from this introgression experiment clearly show that there are some bottlenecks that can be encountered in such an introgression program. These bottlenecks

are first, the selected animals can fail to survive or to reproduce, that is the case observed on homozygous line for chromosome 17. Second few animals with the desired genotype might be available, this is illustrated in the intercross generation 1 (IC_1) reported in Table 5.2. Third only animals of the same sex might be available mainly during the intercross phase; an example can be found in Table 5.2, L_5 within IC_1. These bottlenecks can substantially delay the introgression program and expand the cost involved. Markel et al. (1997) experienced similar problems when introgressing the *Apoe null* allele to multiple inbred strain genetic backgrounds using selection against the donor's genome to speed up the recovery of the recipient background. They observed that the target number of animals, sometimes, was not attained. In their experiment, it happened that the target animal (the best male, that is, the male carrying the *Apoe null* allele and the least of the donor's background) had low viability or more acutely did not breed at all. In plant breeding, attempts to introgress genes from *Allium fistulosum* into *A. cepa* have failed because of sterility in backcrossed generations. Khrustaleva and Kik (2000) overcome this sterility problem by using *A. roylei* as a bridging species.

Sequential genotyping results in this experiment reinforce the theory developed by Koudandé et al. (2000) allowing more than 50% of saving. The method costs time because it makes use of few markers at a time and therefore, is not suitable for genome scans and gene hunting. However, it can be of great importance when introgressing or selecting by means of genetic markers. Results show that when two linked markers are used subsequently (e.g. D5Mit113 and D5Mit10 or D1Mit60 and D1Mit217 in $L_{1,5}$), the second does not make a big difference with the first, resulting almost in the same number of selected animals. As it would be expected, an animal heterozygous at a marker locus is more likely to be heterozygous at a linked marker locus than at other locus located on a different chromosome. When selecting on multiple marker loci, it is useful to apply sequential genotyping on a series of markers that are unlinked, followed by the application of DNA microarrays (Greenfield 2000) for the remaining markers. Such procedure will combine the advantage of sequential genotyping (decrease the total number of genotyping) with the features of DNA microarrays (time saving).

Groups 3, 6 and 9, as indicated in Table 5.2, resulted from the backcross of synthetic mice from groups 1, 4 and 7 with the donor parental line C57BL/6. The LIFETEST results showed that groups 3, 6 and 9 (Table 5.5 and Figure 5.1) were more resistant than the pure donor mice, demonstrating their higher ability to survive trypanosome challenge. This result is consistent with non-additive effect which is not shown by our QTL analysis, but may be the result of heterosis on other parts of the genome different from our three QTL regions. Apart from the introgressed QTL region, the background genotype is expected to carry 3.1% of the donor's genomic DNA at the end of the backcross phase (Young and

Tanksley 1989; Stam and Zeven 1981). Assuming that the mouse genome is 1300 cM and the introgressed chromosomal region carries 20 cM at each side of the marker-bracket, the introgressed regions will represent 7, 9 and 7 percent of the total genome respectively for homozygous mice obtained from lines L_1, L_5 and L_{17}. Recall that the length of the marker-bracket is 5 cM on chromosome 1, 18 cM on chromosome 5, and 7 cM on chromosome 17. In total 90, 88 and 90 percent of the total genome is from the recipient (A/J) mice respectively for homozygous mice obtained from lines L_1, L_5 and L_{17} which correspond to current groups 1, 4 and 7. The cross of these groups with the donor line will result therefore in 0.90, 0.88 and 0.90 genome-wide heterozygosity for groups 3, 6 and 9 respectively. These high levels of heterozygosity can explain the origin of the observed heterosis.

QTL on chromosome 17 was maintained in the linear model because Kemp et al. (1997) have shown that it had the largest effect on survival and explained 14.6% of the total variation in F2 cross of C57BL/6 x Balb/c and 9% in F2 cross C57BL/6 x A/J. In this least square analysis, we dropped the two pure lines used as controls from the data because they had extreme values of background genotype (0 and 1) that had dramatic effects on the estimates. For example, including control inbred lines in the analysis resulted in higher estimates for homozygous mice for donor's QTL allele. Meanwhile the regression coefficient for background dropped to minus 18.8.

From our result on QTL effect analysis, the donor's QTL allele at homozygous status confers longer survival time to animals and is consistent with previous studies (Morisson et al. 1978; Kemp et al. 1996, 1997; Clapcott et al. 2000; Iraqi et al. 2000). From the results in Table 5.5, we can conclude that mean survival times of heterozygous individuals are not different from the mean survival times of homozygous mice. We can infer that donor's QTL alleles on chromosomes 1 and 17 in this study, however, tend to behave as recessive QTL, which is in agreement with the results published by Kemp et al. (1997) for the case of chromosome 17. They did not report on the recessive character of the donor's QTL allele on chromosome 1. In addition they found that the donor QTL allele was recessive on chromosome 5 whereas our current results show an additive effect.

In summary, a successful introgression experiment based solely on genetic markers has been completed. The merit of this experiment is that it gives support to theories advocating the use of genetic markers in relation with QTL. Marker-assisted introgression, however, may be confronted to some bottlenecks such as failure of selected animals to survive or to produce offspring, availability of few animals with the desired genotype, and availability of animals from only one sex, mainly during the intercross phase. Current experiment shows that sequential genotyping can reduce by more than 50% the genotyping load. Further this experiment shows that the additive effect of QTL measured in days of

survival is 9.9 on chromosome 1, 6.8 on chromosome 5, and 6.0 on chromosome 17. The donor's QTL alleles tend to act recessively on chromosomes 1 and 17, whereas they acted additively on chromosome 5. The present experiment shows that introgression is feasible in relatively short time with laboratory animals.

Acknowledgements

O. D. Koudandé is thankful to Robert King, Jane Ikanyi, Henrie N'gathuo and Moses Ogugo from the International Livestock Research Institute (ILRI) for their technical support throughout the experiment. The authors are grateful to The Netherlands Foundation for the Advancement of Tropical Research (WOTRO) and ILRI for financial support. ODK acknowledges the leave provided by INRAB (Institut National des Recherches Agricoles du Bénin) to carry out this research.

Reference

Clapcott SJ, Teale AJ, Kemp SJ (2000) Evidence for genomic imprinting of the major QTL controlling susceptibility to trypanosomiasis in mice. *Parasite Immunol* 22, 259-263

Darvasi A, Soller M (1995) Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* 141, 1199-1207

De Vries JN, Wietsma WA, De Vries T (1992) Introgression of leaf blight resistance from *Allium roylei* Stearn into onion (*A. cepa L.*). *Euphytica* 62, 127-133

Doko A, Guedegbe B, Baelmans R, Demey F, N'Diaye A, Pandey VS, Verhulst A (1991) Trypanosomiasis in different breeds of cattle from Benin. *Vet Parasitol* 40, 1-7

Greenfield A (2000) Application of DNA microarrays to the transcriptional analysis of mammalian genomes. *Mamm Genome* 11, 609-613

Groen AF, Smith C (1995) A stochastic simulation study of the efficiency of marker-assisted introgression in livestock. *J Anim Breed Genet* 112, 161-170

Iraqi F, Clapcott SJ, Kumari P, Haley CS, Kemp SJ, Teale AJ (2000) Fine mapping of trypanosomiasis resistance loci in murine advanced intercross lines. *Mamm Genome* 11, 645-648

Kalbfleisch JD, Prentice RL (1980) *The statistical analysis of failure time data*. New-York, John Wiley & Sons, Inc.

Kemp SJ, Darvasi A, Soller M, Teale AJ, (1996) Genetic control of resistance to trypanosomiasis. *Vet Immunol Immunop* 54, 239-243

Kemp SJ, Iraqi F, Darvasi A, Soller M, Teale AJ (1997) Localization of genes controlling resistance to trypanosomiasis in mice. *Nat Genet* 16, 194-196

Khrustaleva LI, Kik C (2000) Introgression of *Allium fistulosum* into *A. cepa* mediated by *A. roylei*. *Theor Appl Genet* 100, 17-26

Koudandé OD, Thomson PC, Van Arendonk JAM (1999) A model for population growth of laboratory animals subjected to marker-assisted introgression: how many animals do we need? *Heredity* 82, 16-24

Koudandé OD, Iraqi F, Thomson PC, Teale AJ, Van Arendonk JAM (2000) Strategies to optimize marker-assisted introgression of multiple QTL. *Mamm Genome* 11, 145-150

Kristjanson PM, Swallow BM, Rowlands GJ, Kruska RL, De Leeuw PN (1999) Measuring the costs of African animal trypanosomiasis, the potential benefits of control and returns to research. *Agr Syst* 59, 79-98

Lim K-B, Chung J-D, Van Kronenburg CE, Ramana MS, De Jong JH, Van Tuyl JM (2000) Introgression of *Lilium rubellum* Baker chromosomes into *L. longiflorym* Thunb.: a genome painting study of the F_1 hybrid, BC_1 and BC_2 progenies. *Chromosome Res* 8, 119-125

Markel P, Shu P, Ebeling C, Carlson GA, Nagle DL, Smutko JS, Moore KJ (1997) Theoretical and empirical issues for marker-assisted breeding of congenic mouse strains. *Nat Genet* 17, 280-283

Masake RA, Musoke AJ, Nantulya VM (1983) Specific antibody responses to the variable surface glycoproteins of *Trypanosoma congolense* in infected cattle. *Parasite Immunol* 5, 345-355

Morrison WI, Roelants GE, Mayor-Withey KS, Murray WI (1978) Susceptibility of inbred strains of mice to *Trypanosoma congolense*: correlation with changes in spleen lymphocyte populations. *Clin Exp Immunol* 32, 25-40

Murray M, Trail JCM, Davis CE, Black SJ, (1984) Genetic resistance to African trypanosomiasis. *J Infect Dis* 149, 311-319

Murray M, Trail JCM (1984) Genetic resistance to animal trypanosomiasis in Africa. *Prev Vet Med* 2, 541-551

Oertel C, Matzk F (1999) Introgression of crown rust resistance for *Festuca spp.* into *Lolium multiflorum*. *Plant Breeding* 118, 491-496

Roberts CJ, Gray AJ (1973) Studies on trypanosome-resistant cattle II. The effects of trypanosomiasis on N'Dama, Muturu and Zebu cattle. *Trop Anim Health Pro* 5, 220-233

Roelants GE (1986) Natural resistance to African trypanosomiasis. *Parasite Immunol* 8, 1-10

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: A Laboratory Manual*. 2nd ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press)

SAS® User's Guide: Statistics, Version 6.0, 4th Edit. (1990). SAS Inst., Inc., Cary, NC

Soller M, Plotkin-Hazan J (1977) The use marker alleles for the introgression of linked quantitative alleles. *Theor Appl Genet* 51, 133-137

Stam P, Zeven AC (1981) The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. *Euphytica* 30, 227-238

Teale AJ (1993) Improving control of livestock disease. Animal biotechnology in the Consultative Group for International Agricultural Research. *Bioscience* 43, 475-483

Teale A, Agaba M, Clapcott S, Gelhaus A, Haley C, Hanotte O, Horstmann R, Iraqi F, Kemp S, Nilsson P, Schwerin M, Sekikawa K, Soller M, Sugimoto Y, Womack J (1999) Resistance to trypanosomiasis: of markers, genes and mechanisms. *Arch Tierz Dummerstorf* 42, 36-41

Trail JCM, d'Ieteren GDM, Teale AJ, (1989) Trypanotolerance and the value of conserving livestock genetic resources. *Genome* 31, 805-812

Van der Waaij EH, Van Arendonk JAM (2000) Introgression of genes responsible for disease resistance in a cattle population selected for production: genetic and economic consequences. *Anim Sci* 70, 207-220

Yancovich A, Levin I, Cahanar A, Hillel J (1996) Introgression of the avian naked neck gene assisted by DNA fingerprints. *Anim Genet* 27, 149-155

Young ND, Tanksley SD (1989) RFLP analysis and the size of chromosomal segments retained around the *Tm-2* locus in tomato during backcross breeding. *Theor Appl Genet* 77, 353-359

Annex

Derivation of the expected frequencies for genotype-combinations in BC_4

In BC_4 we selected against the C57BL/6 alleles on the markers-QTL that were not the target for the selected line. For example, from line $L_{1,5}$, i.e. heterozygous mice on chromosomes 1 and 5, we wanted to select offspring which were heterozygous for chromosome 1 and homozygous for the donor allele on chromosome 5 and 17, i.e. L_1, therefore, we selected against the donor allele on chromosomes 5 and 17. This meant that we were selecting for the A/J alleles. As no selection had been previously done at chromosomes 17 level during the backcross phase, the proportion of A/J alleles would have increased while that of C57BL/6 would have decreased. The following derivation intends accounting for this variation in the expected frequencies.

During the backcross phase, for example in line $L_{1,5}$, selected individuals were heterozygous on chromosomes 1 and 5, and selected proportion was $p = \left(\frac{1}{2}\right)^z \prod_{i=1}^{z} e^{-d_i}$ where z was the number of unlinked chromosomal regions and d_i the size of each chromosomal region (Koudandé et al. 1999). In other terms p was the product of p_1, p_5 and p_{17} respectively for chromosome 1, 5, and 17 in BC_4 where selection was also planed on chromosome 17. If we were to select for the A/J alleles in BC_1 on chromosome 17, the proportion would be 0.4661 corresponding to p_{17}. Without selection for chromosome 17, the proportion of A/J contribution would increase each generation as follow:

BC_1	$\frac{1}{2}e^{-d} = x$	= 0.4661 (d = 0.07 Morgan).	
BC_2	$\frac{1}{2}(x+1)$	$= \frac{1}{2}(0.4661+1)$	= 0.733
BC_3	$\frac{1}{4}(x+3)$	$= \frac{1}{2}(0.73305+1)$	= 0.866
BC_4	$\frac{1}{8}(x+7)$	$= \frac{1}{2}(0.866525+1)$	= 0.933

The L_1 frequency in BC_4 was then $p_1 \times p_5 \times p_{17}$ = 0.4756*0.4176*0.9332 = 0.1853

A shorter way would be to consider the heterozygosity in BC_3 and subsequently calculate the expected frequency in BC_4. Assuming that selection on the other two chromosomal regions did not affect the allele frequency on the target region, heterozygosity was $\frac{1}{8}$ in BC_3 while homozygosity for A/J alleles was $\frac{7}{8}$. It was expected that $\frac{1}{8} \times p_{17}$ would be

non-recombinant A/J which would result in the expected total frequency of $1/8x + 7/8$, that is the same expression as in previous derivation at BC_4.

Considering the second line $L_{5,17}$, the same expression can be applied, x taking the value of 0.4756 corresponding to $d = 0.05$ Morgan. The contribution of A/J to chromosome 1 was then 0.934 in BC_4. The L_{17} frequency in BC_4 was thus 0.9344*0.4176*0.4661 = 0.1819.

L_5 was selected from the two lines $L_{1,5}$ and $L_{5,17}$. Given M and N the respective numbers of genotyped animals in each line, and assume that f_M and f_N are the frequencies, the frequency of L_5 is calculated as:

$$f_{L_5} = \frac{(M \times f_M) + (N \times f_N)}{M + N} = 0.1841$$

CHAPTER VI

General discussion

Outline

This chapter consists of four sections. In the first section, we discuss the contribution of chapters two through five to introgression in animals using a backcross-breeding program. In the second section, we explore how introgression can be applied in cattle using the current reproduction technologies. The third section describes the environments for which the present study has been designed, i.e. the livestock system in sub-Saharan Africa with an emphasis on the breeding system. Then the fourth section describes constraints limiting introgression, and prospects for the use of trypanotolerance genes in sub-Saharan Africa are discussed.

Genes introgression using genetic markers

Introgression of genes requires their identification and location on the genome. In terms of physiological constituents of an organism such as circulating proteins or enzymes, the corresponding gene can be identified through the generation of mRNA followed by cDNA. Once the cDNA is generated, a fluorescent in situ hybridization on genomic DNA may reveal the site of the involved gene. Conversely in terms of production traits such as growth rate, milk yield and here trypanotolerance, a linkage analysis is performed to relate the traits to some genomic sites called quantitative trait loci (QTL) or sometimes, chromosomal regions. The next step is to locate more accurately the position of the QTL on the chromosome known as fine mapping. Sequencing the chromosomal region and utilizing comparative map information might lead to the identification of the gene(s) involved (Giuffra et al. 2000; Milan et al. 2000).

Fine mapping QTL has been largely used in laboratory animals in the process of identifying genes responsible for traits of commercial importance, therefore many models were proposed to achieve this goal. Inbred lines in laboratory animals have played a key role in identifying and understanding gene and QTL effects. In general, the size of BC and F_2 populations does not allow fine mapping of a gene, as a consequence, the confidence interval of its position is generally large (e.g., 34-140 cM for growth rate QTL in F_2 cross in pigs (Walling et al. 2000)). Advanced generations aim at increasing the number of recombination events and therefore enable a high resolution in mapping. The availability of genotyping data on an advanced intercross line at the ninth generation of a cross between two inbred lines gave an opportunity to evaluate the proportion of recombinant haplotypes in that generation (Chapter 2).

In Chapter 2, the proportion of recombinant haplotypes are quantified using three different methods, a direct counting method, a maximum likelihood (MLE) method, and a

computer software program CRIMAP. From the proportion of recombinant haplotypes, the expected width of interval in which a QTL maps can be predicted. On *Tir1* i.e. trypanotolerance QTL on chromosome 17 (Kemp et al. 1997), results have shown higher proportion of recombinants than expected according to the evaluation of Darvasi and Soller (1995). These results suggest that using F_9 cross will give higher map resolution for trypanotolerant QTL on chromosomes 1, 5 and 17 than the current confidence intervals obtained by Iraqi et al. (2000) on F_6 cross. A higher accuracy of QTL location will result in better monitoring the QTL through selection or introgression. For example, in marker-assisted introgression, one of the advantages of the enhanced precision is that the total number of animals to handle can be kept down as well as the related costs as shown in chapter 3. The major histocompatibility complex (MHC) lies within the 7-cM introgressed QTL region on chromosome 17 in chapter 5, but Kemp et al. (1997) found that the MHC was distinct from the QTL. So, a shorter introgressed QTL region could have prevented the simultaneous introgression of the MHC allele.

Once QTL are detected and located, the next objective is to use this information in practice for improving the production in concerned species (plants or animals) through breeding strategies like marker-assisted selection or marker-assisted introgression (MAI). To use the breeding strategies, there is a need to plan activities, budget and evaluate whether there is a good chance to reach the objectives of such a program. In the particular case investigated in this thesis, i.e. MAI, identification and evaluation of factors influencing the introgression experiment are in chapter 3. When multiple QTL are targeted, major factors influencing the size and cost of the experiment are the number of founder animals, the target number of animals at the end of the backcross phase, the number of backcross generations, and the number and size of QTL regions. Finally several biological factors such as number of offspring per female, mating ratio and sex ratio must be specified. Chapter 3 provides methods to calculate the mean number of animals to handle at each generation through the introgression process, allowing researchers to establish a proper planning and budget for the program. When more than one QTL are to be introgressed, the number of animals to be handled increases tremendously and, therefore, parameters that can be changed to alleviate costs need to be explored. The risk of obtaining fewer numbers of animals than predicted is investigated as well. In animal breeding, the limited reproductive capacity of males and female imposes one to select a large number of animals, which is in marked contrast with some plant breeds in which one seed is enough to generate a population. In contrast, the selection of large number of animals contributes to limit inbreeding in the population. In QTL detection, costs and genotyping load have led to the method of selective genotyping giving approximately the same power of detection as genotyping the whole population (Darvasi and Soller 1992;

Darvasi 1997). This is only advisable if phenotypic information is easy to collect and not expensive as genotyping. When phenotyping endangers animals' life, as in our case here, selective genotyping is not helpful.

The high number of animals needed during the introgression program in mouse led to search for alternatives to minimize both number of animals and amount of genotyping. In the present case of three unlinked QTL, the introgression strategies described in chapter 4 are useful as a model and can also be used in a marker-assisted selection program. This chapter describes methods to optimize a design, estimates the costs involved and chooses the best option accordingly. In this way it can help planning, monitoring and evaluating a project involving marker-assisted selection or introgression. One of the designed strategies has been used in the MAI experiment reported in chapter 5.

As stated earlier, introgression consists of moving a QTL or gene from a certain background genotype to another. Regarding this definition of introgression, our experiment has been successful in mice. The strong points of the MAI experiment reported in Chapter 5 are first, QTL can be controlled by means of linked genetic markers, second, phenotyping is not essential during introgression. Despite this success, there are several points to be stressed:

- It is important to evaluate possible interactions of the QTL to be introgressed, with the original and new background genotype.
- From the introgression experiment in mice as reported in Chapter 5, offspring produced from homozygous individuals for donor's QTL allele on chromosome 17 failed to survive after weaning. The donor's QTL allele may have a negative effect on survival after weaning. Another plausible explanation is that the QTL region carries also some deleterious genes that are originally tempered by some other genes present in their background genotype, genes that are missing in the new background genotype.
- The MAI experiment reveals some bottlenecks that are important to keep in mind for future project. For example, it is not possible to pursue the program when only males or females are available at any step between the last backcross and the last intercross step.

So far, we have identified factors influencing an introgression program, optimized designs for introgression programs, and performed an introgression using laboratory mice. The target animal species, however, is cattle for which the present investigations have been undertaken. From results in mice, we may affirm that QTL introgression using genetic markers is technically feasible in cattle. An introgression program in cattle should be flexible to account for encountered problems with mice.

Marker-Assisted Introgression in cattle

In this section, we discuss some general aspects of introgression in cattle using the results obtained in this thesis. This discussion focuses on the contribution of reproduction technologies to such a program. We successively consider natural service, artificial insemination (AI), a combination of AI and multiple ovulation and embryo transfer (MOET), and other techniques such as in-vitro maturation (IVM), in-vitro fertilization (IVF) and semen/embryo sexing. We close the section by indicating some consequences of introgression for genetic diversity. It is useful, however, to first list some assumptions and conditions.

Assumptions and conditions

We consider three unlinked QTL regions to be introgressed as in Chapter 3. QTL regions are assumed of equal size with a length of 40 cM. Further we assume the sex ratio to be half and the QTL to act recessively. Along this section, we use the introgression strategy (IS_3) described in Chapter 4, which consists in selecting two lines from BC_1 onwards, each line carrying two QTL regions through the backcross steps. The introgression conditions are first, to select males from BC_1 up to the second last backcross (Markel et al. 1997) exploiting their high reproductive capacity. Second, three backcross generations are applied as proposed by Van der Waaij and Van Arendonk (2000) as an economically sound optimum strategy in cattle. We also assume that the maintenance cost for cattle is US$ 2300 (over the time they are in use in the introgression program) and the cost of founder parents is US$ 2500 per head when they are 15 months old. We target 100 homozygous animals at the end of the introgression program. All estimates in this section are based on derivations set up in Chapters 3 and 4.

Natural service

For natural service, assuming that there are enough males and females to complete an introgression program, the total length of the introgression lasts at least 14 years when the number of offspring per female (λ) is one and that the program started when the founder parents were 15 months-old.

Natural service is similar to the mating system used in mice and reported in Chapters 3, 4 and 5, therefore, some conclusions do also apply to cattle. For example with a mating ratio of 20 females per male, the total number of animals to be maintained to achieve 100 homozygous cattle at the end of the introgression is nine million when $\lambda = 1$ and 815 thousand when $\lambda = 2$ (Table 6.1). From chapter 4, the IS_3 strategy is the optimum

compared to the other strategies in terms of costs, number of animals to be genotyped, and number of genotypings when the number of females (r) per male is larger or equal to 3 and if λ is larger or equal to 4. With $\lambda = 4$, the total number of animals drops to 97 thousands in which case the whole introgression lasts at least 32 years. There is a direct relationship between the number of offspring per female and the total number of animals to maintain during the introgression. The reduction of the number of backcross generations does substantially improve this picture. For example, when reducing the backcross phase by one generation, the total number of animals to maintain becomes 65 thousands and costs are reduced by 33%.

There is no reason to believe that genotyping cost in cattle will be different from the result obtained with mice in Chapter 4, however, the DNA sampling cost in cattle might be slightly higher than in mice. With DNA being extracted from hair, we may say that this difference can be ignored. Major costs involved are husbandry costs, which amount US$ 223 millions for the case $\lambda = 4$. There is a possibility to incorporate the husbandry cost in an existing breeding or production program, and in this way keep the introgression cost at reasonable level (Darvasi and Soller 1994; Visscher and Haley 1999). Even for the best case here (i.e. $\lambda = 4$), costs and length of the introgression are not attractive (Table 6.1). In conclusion, an introgression in cattle targeting three QTL is not realistic using natural service.

Table 6.1 Minimal number of cattle, cost and duration of introgression program according to reproductive techniques and number of offspring per female

Number of offspring per female	Factors	Natural service	Artificial insemination	AI + MOET
1	No. of animals	9×10^6	4.6×10^6	-
	Duration (years)	14	14	-
2	No. of animals	8.15×10^5	6.0×10^5	-
	Duration (years)	20	20	-
4	No. of animals	9.7×10^4	9.0×10^4	-
	Duration (years)	32	32	-
	Cost (US$)	2.23×10^8	2.07×10^8	-
25	No. of animals	-	-	1.6×10^3
	Duration (years)	-	-	14
	Cost (US$)	-	-	3.9×10^6

MOET = multiple ovulation and embryo-transfer.

Artificial insemination

With AI, it is sufficient to select only one male at each backcross generation, which will result in less animals to be produced and less genotyping to be performed compared to natural service. The total duration of the introgression, however, does not differ from the case of natural service, i.e. at least 14 years when $\lambda = 1$ or 20 years when $\lambda = 2$.

The expected impact of AI on the total number of animals to handle is that it will reduce this number and consequently reduce the related costs. For example the drop in this number (4.6 millions) is about 50% of the corresponding number in natural service when $\lambda = 1$. In case $\lambda = 4$, the husbandry cost is about 207 millions (Table 6.1). Costs are still high even though insemination costs are ignored in this calculation. It is also expected that, as in the case of natural service, incorporating husbandry costs in an existing breeding or production program will help keeping the introgression costs at a reasonable level.

Selecting one male at each backcross generation has the disadvantage of creating a population with high level of inbreeding at the crossing and intercrossing phase of the introgression. One way to limit inbreeding in the generated population is to increase the number of sires used across the backcross phase in which one bull is selected within a sire's half sib family. This will result in higher numbers of animals to handle than in case of selecting one male at each backcross generation.

Whether it is a natural service or artificial insemination, the number of females carrying the desired chromosomal regions limits the crossing and intercrossing phases. Indeed, at this step of the introgression, both males and females should be carrying the desired QTL alleles and the limited females' reproductive capacity becomes critical for the process. This limitation can be reduced when using multiple ovulation and embryo transfer (MOET).

AI combined with MOET in an introgression program in cattle

Although MOET (Nicholas and Smith 1983) can be used in the backcross phase (mainly when the recipient breed is of a high value), its use becomes more important during later phases (i.e. lines cross and intercross) where the females' reproductive capacity is a limiting factor. MOET can help in the way that it can provide many more embryos than a female can produce offspring and that in a shorter time. Suggestion to use MOET in the later phases of the introgression is reinforced by the design where only males carrying the donor's QTL allele are used during the backcross phase. Cunningham (1999) stated that with conventional embryo transfer, a cow could produce as many as 25 calves in her lifetime. Recipient cows, other than embryos' donors, are needed to carry out embryos development, and therefore, should be accounted for in the introgression evaluation.

Keeping the number of calves per female at 25, the total number of animals to maintain during the introgression is 1600 (Table 6.1). Husbandry costs are estimated at US$ 3.7 million while embryo transfer costs are estimated at more than US$ 100 per unit (Cunningham 1999). Applying US$ 150 per unit, MOET costs US$ 186150 resulting in a total of US$ 3.9 million for the introgression program. This results is encouraging compared to previous reproduction systems whose costs were hundreds of millions US dollars. However the level of inbreeding will be somewhat higher than when only AI is used. Using MOET increases the number of related animals and, therefore, will increase the inbreeding level in the final population.

The total length of the introgression using AI and MOET as described here is 14 years corresponding to the shortest time of realizing such a program. Improving MOET by increasing the number of calves per females will not shorten the introgression process, but will decrease the total number of animals to maintain. Other techniques such as in vitro maturation (IVM) and in vitro fertilization (IVF) can shorten the introgression process.

Other techniques

IVM and IVF are new techniques in which immature oocytes are collected from living animals or from slaughterhouse, matured and then fertilized in vitro (Gordon 1994; Tervit 1996). These techniques can improve the efficiency of an introgression program, for example by enhancing the number of offspring per female. Other feature of IVM is that it can contribute to shorten the length of introgression programs in the way it is no longer necessary to wait till 15 months age of the young female. Oocytes can be aspired from pre-pubertal females, then matured and fertilized in vitro in laboratory.

Embryo sexing is a technique still under investigations and not widely used. It consists in identifying the sex of an embryo by using an appropriate DNA probe on one cell extracted from the embryo (Thibier and Nibart 1995). There is a concern about the impairment that the sampling can cause to the embryo and the consequence on its viability and fertility. With progress in technology, we can expect that marker genotyping will be feasible at embryo level without jeopardizing its viability. In this condition, selection can early take place, and only embryos carrying desired alleles will be implanted.

Determining whether the sperm used for fertilization is carrying an X or a Y chromosome indicates the sex of the resulting embryo. Current technology for sperm sexing uses the difference between sperms carrying X and Y chromosome and, by means of a laser beam, the sperm is sorted into male or female (Cran and Johnson 1996) using the standard flow cytometry equipment. The efficacy of the system is 90 percent correct

sorting but the sperm can be damaged during the process resulting in a reduction of the average fertility.

Overall, these techniques can be beneficial for an introgression program as long as they are cost effective. Indeed we have seen that a type of sex is desired at certain step of the introgression and these techniques can provide the program with embryos of the desired sex.

Genetic diversity

Introgressing trypanotolerance QTL in other cattle breeds, one should care about the *donor* breed if the new trypanotolerant cattle breed was to be used in the same environment. In that case, there is a risk that the new breed, potentially high producer, may progressively replace the *donor* cattle breed leading to its extinction. A conservation program would be appropriate to preserve the *donor* breed.

One of the main goals of introgression is to recover the *recipient* genome as much as possible so as to approach its original production level. One way of recovering the *recipient* gene pool is to increase the number of backcross generations. To speed up this recovery process, one can apply selection against the donor background genotype (Markel et al. 1997; Wakeland et al. 1997). Restoration of the *recipient* gene pool can be achieved except at the introgressed QTL region. In theory, in laboratory animals (e.g., mice, drosophila) and in plants, one can virtually recover the gene pool, but this will be very difficult to achieve in farm animals. The genetic diversity in the new cattle breed might be consistent because it combines the *donor* as well as the *recipient* gene pool.

Livestock production environment in sub-Saharan Africa

The introgression program as described in previous section implicitly involves intensive cattle production systems corresponding to what can be met in developed countries. Trypanosomosis, however, is a concern of developing countries especially in sub-Saharan Africa where the disease impaired the development of cattle production, therefore, current and next sections will stress on particularities related to this part of Africa as far as introgression is concerned.

The distribution of ruminants, mainly cattle, and the production systems in sub-Saharan Africa are strongly dependent on agro-ecological conditions. Five major agro-ecological zones are distinguished in sub-Saharan Africa, that is, the arid, semi-arid, sub-humid, humid and highland zones (Figure 6.1).

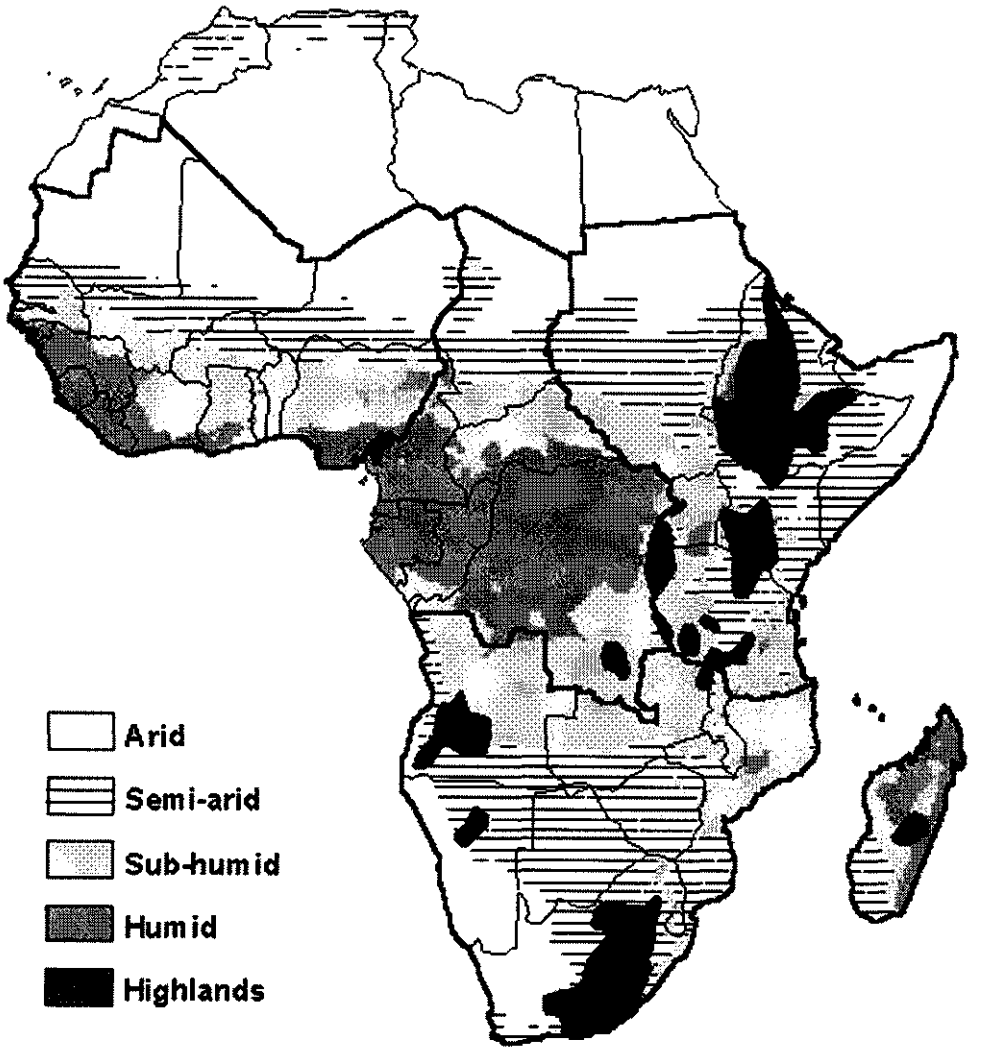

Figure 6.1 Agro-ecological zones in sub-Saharan Africa

The arid zone is characterized by a variable rainfall that precludes cropping most of the time, except in oases or areas under irrigation. The main vegetation types are short annual grasses that loose their freshness just at the end of the rainy season that is pretty short. Serious degradation of land has occurred around water points and areas of permanent human habitation. Traditional nomadic and transhumant pastoral systems based upon communal grazing prevail in the arid zone (Winrock 1992; De Haan et al. 1996). The

systems are well adapted to the sharp annual and seasonal variation in rainfall, requiring mobility among pastoralists to move where forage may be found. Traditional pastoral systems make efficient use of the vegetative resources of the zone, but the rangelands are stocked near its limits and there is little potential for increasing off-take (Winrock 1992). Breeding is performed by natural service where males are permanently kept in the flock.

The semi-arid zone covers one fourth of the sub-Saharan Africa and population densities can be as high as 250-300 inhabitants per square kilometer in the main West African River valleys. Livestock exist in nomadic and transhumant systems in the driest areas and provide much of the value of agricultural output. Great interaction with crop farming and large scale intrusion of cropping into the rangelands are observed in this part of the semi-arid zone. In the higher rainfall areas, livestock are raised mainly as components of smallholder mixed crop-livestock systems. The advantage of this zone for livestock production is the absence of trypanosomosis and the availability of abundant pasture of good quality. With the increase in population pressure, large part of the available good pasture is being converted into croplands (De Haan et al. 1996). Twenty five percent of sub-Saharan Africa's ruminant livestock is in the semi-arid zone with 50 million cattle, 28 million sheep, and 38 million goats (Winrock 1992). In the semi-arid zones, declining soil fertility, inadequate dry season feed, scarce water, are constraints for agriculture driven development. Due to the expansion of cultivation, there is a threat on rangelands' productivity and on pastoral production systems because crop cultivation gives rights for land use in traditional systems, that is not the case for grazing. The increasing use of livestock as savings (mainly by people living in urban areas) favors more the number of heads than productivity and contributes to overload the capacity of the region.

The sub-humid zone is characterized by less variable rainfall than the arid and semi-arid zones, offering better conditions for crop and pasture production. Pastoralists from the drier areas are increasingly moving into the sub-humid zone for dry season grazing which results in serious conflicts when seasonal movement occurs earlier than crop harvest takes place. Vegetation in this zone is more susceptible to degradation as a result of mismanagement than the predominantly annual vegetation of the arid zones. Human population density as well as livestock density is low because of high risk of disease for livestock (including trypanosomosis), but the situation is evolving with increasing human population. This increase alters the zone's ecology resulting in a decrease of tsetse flies and subsequently to a lesser risk of trypanosomosis. The potential capacity of the sub-humid zone to carry ruminant livestock is exploited at about 18 to 37% of its capacity (Winrock 1992). The domestic ruminant population consists of 37 million cattle, 18 million sheep and 24 million goats. Soils and consequently forages are of poor quality and during the dry season, protein contents of mature forages drop below 5 percent (Mohamed

Saleem and von Kaufmann 1991). The fragile soils, poor quality forages, diseases pressure, shortage of infrastructure for transportation, lack of motivation for an economical exploitation of livestock (FAO/ILRI 1995) are limiting factors for livestock development in the sub-humid zone.

The humid zone consists of rain forests and derived savannas, and is poorly inhabited except in eastern Nigeria and around big cities. The number of ruminant livestock is low, 10 million cattle, 10 million sheep, and 14 million goats. Forages are of low quality and the major limiting factor for livestock production is trypanosomosis (Winrock 1992). Intensive commercial poultry and pig production have evolved in the suburban areas. De Haan et al. (1996) found that there was a strong shift in livestock production from the arid and highland zones to more humid zones following recent advances in disease control technologies. They predicted that livestock production is likely to become more crop-based.

Covering only 5 percent of the sub-Saharan Africa, highland zones have the greatest density of people (15 percent) and livestock. The ruminant population is 32 million cattle, 26 million sheep, and 14 million goats. This zone is a net exporter of meat and living animals to other agro-ecological zones and small net importer of milk. The most common farming system is smallholder crop-livestock farms. Animal traction is differently used depending on regions, cropping patterns and farm size. Because of a favorable climate, moderate disease and pest problems, and a high production potential, the zone is attractive for people and constitutes a good environment for livestock. Diseases, however, can be a severe problem in areas where some microclimates can favor their expression.

Environmentally independent animal production exists and is expanding. Intensive commercial systems rely heavily on investment in technology and inputs and tend to be little affected by agro-ecological conditions, particularly for poultry, pigs and for dairy production in suburban areas. These systems have evolved following the increased demand for meat, milk and eggs, and are much more market-oriented than the grassland/roughage based production system.

Prospects for the use of trypanotolerance genes in Africa

Demands for animal products

It is reported that imported milk represents a large part of the commercial exchanges in Africa economics (Winrock 1992). However, given the price of this milk in developing countries, few rural people can afford it and almost all this imported milk is consumed in urban areas. Further, the World Health Organization recommendation of breast-feeding for

babyhood contributes to the decreasing demand on imported milk. In addition, milk is not part of the diet for some populations in sub-Saharan Africa. Nevertheless, increasing human populations in cities coupled with a better buying power constitutes a potential market for milk production. An improved buying power will stimulate a higher demand for food quality of which milk holds an outstanding position and will become an important driving force for a breeding program aiming to improve livestock.

There are alternatives for producing valuable proteins for human consumption. For example, promoting eggs production in rural areas can make a cheap source of animal protein available. At current stage, meat production is deficient because livestock production is not market-oriented. A change in the perception of livestock from a contemplative to an economical issue will trigger the development of this agricultural sector in sub-Saharan Africa. Such change, for example, has been observed in Sahelian countries (Burkina Faso, Mali) where the off-take on livestock has increased after a devaluation of the currency FCFA[1] in 1994 (De Haan et al. 1996) resulting in less pressure on the rangelands. Further, with the increasing population in Africa, farming systems tend to evolve and become more intensive than previously observed. For example in south of Bénin and Nigeria, the increasing human population resulted in shortage of plots for cropping. Subsequently, farmers started cropping around their villages, and as a result, small ruminants production system evolved from free roaming to pen management.

Trypanosomosis burden

Trypanosomosis is the most widespread disease of great importance in Africa affecting humans and animals. It has a dramatic impact on agriculture: the disease in human restricts labor involvement in agricultural production, prevents the use of lands colonized by tsetse flies for agriculture activities. In farm animals, trypanosomosis impedes production and may be fatal for the animals if not treated. No new drugs are being developed and an increasing resistance to few available drugs is observed on trypanosomes. Vaccine is not available and attempts to eradicate tsetse flies from the environment have not been successful. The disease costs approximately US $1340 million per year for livestock producers and consumers in Africa (Kristjanson et al. 1999). For all these reasons, the well adapted indigenous cattle breeds (N'Dama, Baoulé, Lagune, Muturu) are promising genetic resources for livestock production.

Experts in agriculture reported that, due to trypanosomosis, humid and sub-humid zones of Africa (corresponding to tsetse-infested zones, see Fig. 1.1 on page 4) are under-exploited at about 18 to 37% of its current capacity (Winrock 1992). It is estimated that

[1] Franc de la Communauté Financière Africaine

approximately 7 millions square km could be used for livestock or mixed farming without stress to the environment if trypanosomosis was controlled. It is then suggested that introducing some local or synthetic trypanotolerant cattle breeds would help improving animal production. A cattle breed, which is tolerant due to the introgression of trypanotolerance genes, gains ability to live in the sub-humid and humid zones of Africa. The immediate consequence is an increase in cattle as well as human populations in these zones. As a result, there will be a higher demand on the existing and fragile resources that can lead to soil degradation. In addition, the introduction of trypanotolerant cattle in humid and sub-humid zones can evolve competition for habitat and feed resources between the exotic (synthetic trypanotolerant) cattle and wildlife. Recently in East-Africa, Voeten (1999) showed that cattle populations share the same feed resource with zebra and wildebeest in wet season whereas no overlapping was observed in the dry season. For all these reasons, trypanotolerant breeds should be introduced with care in the sub-humid and humid zones. Accompanying measures should be implemented wherever a trypanotolerant breed is produced or introduced to avoid damage to local environment. An improved management should target the availability of water supply (possibility to exploit rainfall) and the settlement of pastures.

Breeding for the use of trypanotolerance genes

An option for using trypanotolerance gene is to select within trypanotolerant breeds for production traits. In this particular case of N'Dama, it might be beneficial to select for meat and milk traits (Van der Waaij and Van Arendonk 2000) and the program should be executed in the local and harsh environment. The implementation of nucleus breeding schemes, from which the genetic gain will flow into the surrounding flocks through the top bulls in the nucleus, would be appropriate. At the beginning of the program, the use of a simple breeding scheme as suggested by Smith (1988) for selection and improvement at subsistence level should be the priority.

Trypanotolerance genes can also be used by exploiting F1 crossbred animals produced by crossing European dairy cattle with local trypanotolerant cattle. This scenario is already in use in a large number of countries in Africa. The arising problem is the furniture of European cattle breeds, which is currently done through semen importation.

Trypanotolerance genes introgression may be a good alternative to improve the resistance to trypanosomosis of exotic high milk producing breeds imported in Africa. Such introgression could be implemented where these exotic cattle breeds already exist, i.e. suburban areas in risky zones. Highland zones may also be used to produce these introgressed animals with trypanotolerant genes, and exported in regions where they are

needed. These production systems (dairy production) are market-oriented and might be more receptive to an introgression program. Trypanotolerance genes introgression may also be a good alternative to enhance the resistance of local zebu breeds to trypanosomosis, and in that way enables beef production in tsetse-infested zones. The most likely and suitable scheme to develop this introgression is an open nucleus breeding (Cunningham 1980). It will consist of specific introgression centers created at different points in tsetse-infested as well as tsetse-free areas. Depending on the sensitivity of the recipient animals to tsetse challenge, they will be assigned to one of these two areas. Reproductive techniques such as artificial insemination and embryo-transfer should be methods of choice. Farmers willing to participate will be involved in the introgression program. The results, however, will be available for all farmers.

An alternative to trypanotolerance gene introgression is to introgress some milk production genes into well-established trypanotolerant cattle breeds in Africa to improve milk production. This alternative is possible as long as the number of genes to introgress is not too high. Resulted synthetic animals will have better chance to maintain their other features such as ability to cope with heat stress, adaptation to feeding on low-nutritive value stuff and adaptation to other intercurrent diseases common in sub-Saharan Africa.

Implementation of Marker-Assisted Introgression

As seen in section 2 of this chapter, introgression requires a large amount of resources, and animals involved should be monitored consistently throughout that process. In section 3, the prevailing production systems in sub-Saharan Africa (rangelands, nomadism and transhumance), climatic, nutritional and health conditions, and the absence of structured breeding programs are constraints for using introgression to develop a synthetic breed. Lack of adequate infrastructure (roads, paths) and means of transportation due to poor economics, together with small cattle populations makes the cost per insemination high. Also blurred heat manifestations in indigenous cattle are reasons that render AI costly and difficult if not impossible to implement (Bondoc et al. 1989). Deficient recording, inability to process data, and delay or no feedback to the breeders are other constraints (Vaccaro and Steane 1990) that can impede an introgression program. Cryopreservation for semen is still an expensive technology in this part of Africa. Using natural service is not advisable given the multiple problems to face: progeny will spread over a long period due to the limited number of services a male can perform each day, and because heat period might also be scattered over time. In addition, the number of animals and the cost involved are other strong reasons that do not favor an introgression program in the most widespread livestock production systems in Africa. This leaves dairy cattle, mainly in suburban areas,

as the only suitable alternative system to undertake an introgression program targeting trypanotolerant genes. However, with growing population in expanding cities, there is a decreasing tsetse pressure resulting in a decreasing risk for trypanosomosis, therefore, dairy cattle can be used for introgression and then disseminated in tsetse-infested areas.

Priorities of different governments in sub-Saharan Africa are far from creating resistant population of cattle to trypanosomosis. Monitoring drought and flood disasters, hunger, civil war, AIDS and other diseases in human is the most common daily commitment. In these conditions, it is very difficult for governments to allocate resources to breeding programs.

A center like the International Livestock Research Institute (Nairobi-Kenya) might be committed to create the synthetic breed in collaboration with for example, the International Trypanotolerance Center in The Gambia. The next arising question is "who is going to pay for the creation of the synthetic breed?" The end users, i.e. the farmers involved in milk production business as well as small farmers, might be interested in using the results but are likely not interested in sponsoring the program to develop the synthetic breed. This implies that it becomes the governments' duty to sponsor this program through the budget for development program. An international funding assistance might be very helpful given the economic situation of different countries involved. A breeding program for trypanotolerant cattle to produce meat and milk is not just an individual country's business, but a regional responsibility.

Trypanotolerance gene introgression is without doubt a promising future for exploiting large areas in sub-humid and humid zones of Africa where trypanosomosis prevents some cattle breeds to produce and even acutely to survive without drug use. Introgression of trypanotolerance gene does not mean that there will be no use of drugs, but there will be a limitation in the number of treatments required. As mentioned earlier, care should be taken to limit negative impacts on the environment. In addition, one should keep in mind that there are many other important pathogens that can hide the benefit expected from introgression. Performing trypanotolerance gene introgression in tsetse challenged areas may help to measure the real interactions with other diseases.

References

Bondoc OL, Smith C, Gibson JP (1989) A review of breeding strategies for genetic improvement of dairy cattle in developing countries. *CAB International* 57, 819-829

Cran DG, Johnson LA (1996) The predetermination of embryonic sex using flow cytometrically separated X and Y spermatozoa. *Hum Reprod Update* 2, 355-363

Cunningham EP (1980) Methods for recording, evaluation and selection in adverse environments. *FAO/UNEP technical consultation on animal genetic resources, conservation and management, Rome, 12 p*

Cunningham EP (1999) The application of biotechnologies to enhance animal production in different farming systems. *Livest Prod Sci* 58, 1-24

Darvasi A., Soller M (1992) Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theor Appl Genet* 85, 353-359

Darvasi A, Soller M (1994) Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. *Theor Appl Genet* 89, 351-357

Darvasi A, Soller M (1995) Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* 141, 1199-1207

Darvasi A (1997) Interval-specific congenic strains (ISCS): an experimental design for mapping a QTL into a 1-centimorgan interval. *Mamm Genome* 8, 163-167

De Haan C, Steinfeld H, Blackburn H (1996) *Livestock and the environment: finding a balance.* WRENmedia, Suffolk, UK, 115 p

FAO/ILRI (1995) Livestock development strategies for low income countries. *Proceedings of joint FAO/ILRI roundtable on livestock development strategies for low-income countries held at International Livestock Research Institute, Addis Ababa, Ethiopia, 27 February – 2 March 1995*

Giuffra E, Kijas JMH, Amarger V, Carlborg Ö, Jeon J-T, Andersson L (2000) The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* 154, 1785-1791

Gordon I (1994) *Laboratory production of cattle embryos.* CAB International, Wallingford, U.K.

Iraqi F, Clapcott SJ, Kumari P, Haley CS, Kemp SJ, Teale AJ (2000) Fine mapping of trypanosomiasis resistance loci in murine advanced intercross lines. *Mamm Genome* 11, 645-648

Kemp SJ, Iraqi F, Darvasi A, Soller M, Teale AJ (1997) Localization of genes controlling resistance to trypanosomiasis in mice. *Nat Genet* 16, 194-196

Markel P, Shu P, Ebeling C, Carlson GA, Nagle DL, Smutko JS, Moore KJ (1997) Theoretical and empirical issues for marker-assisted breeding of congenic mouse strains. *Nat Genet* 17, 280-283

Milan D, Jeon J-T, Looft C, Amarger V, Robic A, Thelander M, Rogel-Gaillard C, Paul S, Iannuccelli N, Rask L, Ronne H, Lundström K, Reinsch N, Gellin J, Kalm E, Le Roy P, Chardon P, Andersson L (2000) A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. *Science* 288, 1248-1251

Mohamed Saleem MA, von Kaufmann RR (1991) Fodder bank: An innovation for improving the dry season nutrition of traditionally managed cattle in the sub-humid zone of West Africa. Paper presented at Animal Production Global Workshop, September 1991, San Jose, Costa Rica

Nicholas FW, Smith C (1983) Increased rates of genetic change in dairy cattle by embryo transfer and splitting. *Anim Prod* 36, 341-353

Smith C (1988) Genetic improvement of livestock using nucleus breeding units. *World Anim Rev* 65, 2-10

Tervit HR (1996) Laparoscopy/laparotomy oocyte recovery and juvenile breeding. *Anim Reprod Sci* 42, 227-238

Thibier M, Nibart M (1995) The sexing of bovine embryos in the field. *Theriogenology* 43, 71-80

Vaccaro L, Steane DE (1990) Practical technologies and options for genetic improvement of livestock in developing countries. In *Strategies for sustainable animal agriculture in developing countries*. FAO Animal Production and Health Paper 107, 85-98

Van der Waaij EH, Van Arendonk JAM (2000) Introgression of genes responsible for disease resistance in a cattle population selected for production: genetic and economic consequences. *Anim Sci* 70, 207-220

Visscher PM, Haley CS (1999) On the efficiency of marker-assisted introgression. *Anim Sci* 68, 59-68

Voeten MM. (1999) *Living with wildlife: coexistence of wildlife and livestock in an East-African savanna*. Doctoral thesis, ISBN 90-5808-133-8 Wageningen University

Wakeland E, Morel L, Achey K, Yui M, Longmate J (1997) Speed congenics: a classic technique in the fast lane (relatively speaking). *Immnol Today* 18, 472-477

Walling GA, Visscher PM, Andersson L, Rothschild MF, Wang L, Moser G, Groenen MAM, Bidanel J-P, Cepica S, Archibald AL, Geldermann H, de Koning DJ, Milan D, Haley CS (2000) Combined analyses of data from quantitative trait loci mapping studies: chromosome 4 effects on porcine growth and fatness. *Genetics* 155, 1369-1378

Winrock (1992) *Assessment of animal agriculture in Sub-Saharan Africa*. Winrock International, Morrilton, USA, 125p

Summary

Trypanosomosis is the most widespread disease of great importance in Africa affecting humans and animals. The disease called sleeping sickness in human invalids infected population. As a consequence, it restricts labor involvement in production and prevents the use of lands colonized by tsetse flies for agriculture activities. In farm animals, trypanosomosis impedes production and may be fatal for the animals if not treated. No new drugs are being developed and an increasing resistance to few available drugs is observed on trypanosomes. A vaccine is not available yet and attempts to eradicate tsetse flies have not been successful. The disease costs approximately US $1340 million per year for livestock producers and consumers in Africa. For all these reasons, the well adapted indigenous cattle breeds (N'Dama, Baoulé, Lagune, Muturu) are promising genetic resources for livestock production.

Experts in agriculture reported that, due to trypanosomosis, humid and sub-humid zones of Africa, corresponding to tsetse-infested zones, are under-exploited at about 18 to 37% of its current capacity. It is estimated that approximately 7 millions square km could be used for livestock or mixed farming without stress to the environment if trypanosomosis was controlled. It is then suggested that introducing some local or synthetic trypanotolerant cattle breeds would help improving animal production.

A sustainable control option of trypanosomosis involves the farming of trypanotolerant livestock. Experiments to map genes involved in trypanotolerance in N'Dama cattle using a linkage approach are well advanced. Identification of trypanotolerance genes opens the way for marker assisted introgression to incorporate trypanotolerance in other breeds of cattle. The overall aim of this thesis is to develop optimum breeding schemes for the introgression of trypanotolerance genes using mouse as an experimental model organism.

Introgression of genes requires their identification and location on the genome. Advanced intercross generations aim at increasing the number of recombination events and therefore enable a high resolution in mapping. **Chapter 2** assesses the proportion of recombinant haplotypes in an advanced intercross F_9 population of mice (A/J x C57BL/6), based solely on genotyping data collected from that population. The targeted chromosomal region of approximately 7-cM is located on chromosome 17 and contains a major QTL associated with trypanoresistance in mice. The estimation of the proportion of recombinant haplotypes is performed through a counting method and maximum likelihood estimation (MLE) method that does and does not account for misclassifications. A simulation study is performed to evaluate the characteristics of these two methods. Finally CRIMAP program

has been used to assess the same data set. Allele frequencies within the F_9 population show deviations from the expected frequency (0.5); these deviations are likely due to random drift. The counting method gives estimates for the proportion of recombinant haplotypes in F_9 that are biased upward. Results from the maximum likelihood model suggest 63% higher proportion of recombinant haplotypes whereas CRIMAP analysis doubles the value obtained from the mouse genome database. The current data set suggests a different marker order other than that reported by the mouse genome database. Misclassification is an important factor to be considered when analyzing such data even though the accuracy of the estimates of misclassified fraction is low. The results of this study clearly show evidence of a strong accumulation of recombinant haplotypes within the target chromosomal region in the F_9.

Chapter 3 provides methods for calculating the mean and variance of the number of animals with the desired genotype in each backcross generation for a marker assisted introgression experiment. The ultimate goal is to produce animals that are homozygous for the desired loci. The methods have been developed specifically for experiments with inbred lines. The model assumes a Poisson distribution for litter size, and is similar to that used in stochastic versions of population dynamics models. Certain biological parameters have to be specified as well as parameters under the control of the breeder. These methods can be utilized in designing an experiment to determine the number of founder animals required, given the number of animals required at the completion of the backcross process and vice versa. Consideration is also given to minimizing the total amount of genotyping over the entire experiment, by varying the number of times each backcrossed male is used. In addition, an outline is given for an adaptive design that allows for changes in male usage to be made during the experiment.

To optimize designs for marker-assisted introgression programs aimed at introgressing three unlinked quantitative trait loci (QTL), **Chapter 4** studies different alternatives versus a traditional backcross or intercross phase. The alternative backcross strategies appear to be more advantageous by having 50% less genotyping load than a traditional backcross strategy tracking all three QTL at a time through a single line. A multiplication phase following the selection of homozygous animals at the three QTL as an intercross alternative allows doubling of the number of homozygous animals in a mouse model compared with the first intercross generation. Within the same model, a second intercross alternative with individuals carrying all three QTL at the first intercross results in a 12-fold increase in the number of homozygous animals obtained in the first intercross generation. The same ranges of decrease are observed in the number of animals to be genotyped and

the number of genotypings when targeting a fixed number of homozygous animals. An option, with two lines each carrying two QTL through the backcross phase and coupled with the second intercross alternative, appears to be the best introgression alternative. This option requires 76% fewer genotypings, 68% fewer animals to be genotyped, and costs 75% less than an option in which all three QTL are introgressed through a single line.

A marker-assisted introgression (MAI) experiment was conducted to transfer trypanotolerance QTL from a donor mouse strain, C57BL/6, into a recipient mouse strain, A/J. The objective was to assess the effectiveness of such introgression using genetic markers and evaluate the effect of each of the three QTL in the new background genotype. We used a backcross strategy that consisted in selecting two lines, each carrying two of the donor QTL alleles through the backcross (BC) phase. At the fourth BC generation, single carrier animals were selected to proceed with the intercross phase. Given the genotype on chromosomes 1, 5 and 17, ten groups of mice plus two controls (A/J and C57BL/6) were constituted at the end of the intercross phase and went through the challenge with *Trypanosoma congolense*. Sequential genotyping was applied to BC_3 animals. In **Chapter 5**, we report on the results obtained during the first 90 days post-inoculation. The results showed that markers were transmitted from a generation to another and could be traced through the breeding phase. During this phase, some bottlenecks were encountered: failure of selected animals to survive or to reproduce, availability of few animals with the desired genotype, and availability of animals from only one sex which is crucial during the intercross phase. Sequential genotyping reduced the genotyping load by more than 50%. Females had longer mean survival time than males ($P < 0.001$). Significant effects were found for all three QTL. The additive effect of the QTL measured in days of survival was 10.0 on chromosome 1, 7.0 on chromosome 5, and 6.0 on chromosome 17. The proportion of the donor background genotype had a remarkable effect on survival time. The donor's QTL alleles tended to act recessively on chromosomes 1 and 17 whereas alleles acted additively on chromosome 5. The present experiment shows that introgression is feasible and effective in relatively short time with laboratory animals.

In **Chapter 6**, we discuss the contribution of chapters two through five to introgression in animal species through a backcross-breeding program using genetic marker information. Next we explore how introgression can be applied in cattle using the current reproduction technologies. Subsequently, we review the livestock environment in sub-Saharan Africa with an emphasis on the breeding system. The chapter is closed with some prospects for using trypanotolerance genes in sub-Saharan Africa. An option for using trypanotolerance gene is to select within trypanotolerant breeds for some production traits. In this particular

case of N'Dama, it might be beneficial to select for meat and milk traits. Trypanotolerance genes can also be used by exploiting F1 crossbred animals produced by crossing European dairy cattle with local trypanotolerant cattle. This scenario is already in use in a large number of countries in Africa. Introgressing trypanotolerance genes may also be a good alternative to improve the resistance of exotic high milk producing breeds imported in Africa to trypanosomosis. Another alternative for trypanotolerance gene utilization is to introgress some milk production genes into well-established trypanotolerant cattle breeds in Africa to improve milk production. To undertake an introgression program, an international funding assistance might be very helpful given the economic situation of different countries involved. A breeding program for trypanotolerant cattle to produce meat and milk should not be just an individual country's business, but a regional responsibility

Résumé

La trypanosomose est la maladie de grande importance la plus répandue en Afrique affectant à la fois les hommes et les animaux. La maladie invalide les hommes du point de vue de la force de travail et empêche ainsi l'utilisation des terres colonisées par les mouches tsetse pour les activités agricoles. Chez les animaux domestiques, la trypanosomose limite la production et dans les cas extrêmes peut être mortelle si les animaux ne sont pas traités. Aucun nouveau médicament n'est en voie d'être développé et une résistance croissante des trypanosomes est observée vis-à-vis du peu de médicaments disponibles. Il n'y a pas de vaccin disponible et les tentatives d'éradiquer le vecteur ont échoué. La maladie coûte environ 1340 millions de dollars US par an aux éleveurs et consommateurs de bétail en Afrique. Les bovins trypanotolérants d'Afrique de l'Ouest (N'Dama, Baoulé, Lagune, Muturu) constituent donc un potentiel génétique pour la production animale.

Les experts en matière d'agriculture ont estimé qu'en raison de la trypanosomose, les zones humides et sub-humides de l'Afrique, correspondant aux zones de distribution des mouches tsetse, sont exploitées à environ 18 à 37% de leur capacité actuelle. On estime qu'environ 7 millions de kilomètre carré pourraient être utilisés pour l'agriculture sans dommage à l'environnement si la maladie était maîtrisée. Il est donc suggéré d'introduire et d'exploiter les races locales de bétail trypanotolérant dans ces zones infestées de glossines pour aider à améliorer la production animale.

Une option durable de contrôle de la trypanosomose est l'élevage du bétail trypanotolérant. Des expériences en cours sur des croisements de zébu Boran avec N'Dama et utilisant la méthode d'analyse d'association pour la cartographie des gènes de trypanotolérance sont bien avancées. L'identification des gènes liés à la trypanotolérance ouvre ainsi la voie à l'introgression assistée par marqueurs de ces gènes sur les animaux de race sensible. Le but général de cette thèse est de développer des schémas optimaux de croisement en vue de l'introgression des gènes de trypanotolérance utilisant la souris comme modèle d'organisme expérimental.

L'introgression des gènes nécessite d'abord leur identification et localisation sur le génome. Les générations avancées de croisement interne (intercross) visent à augmenter le nombre d'événements de recombinaison sur le génome et par conséquent permettent une cartographie de haute précision. Le **chapitre 2** évalue la proportion d'haplotypes de recombinaison dans une population provenant de la neuvième génération d'intercross entre des souris A/J et C57BL/6, se basant uniquement sur les informations des génotypes de cette population. La région chromosomique visée dans cette étude et mesurant environ 7-

cM est située sur le chromosome 17 et contient un locus de caractère quantitatif (QTL) majeur associé à la trypanorésistance chez les souris. L'estimation de la proportion d'haplotypes de recombinaison est réalisée en utilisant la méthode de comptage et la méthode d'évaluation par maximum de vraisemblance qui prend ou ne prend pas en compte les erreurs de classification des individus. Une étude de simulation est réalisée pour évaluer les caractéristiques de ces deux méthodes. Enfin le programme CRIMAP a été employé pour évaluer les mêmes données. Les fréquences des allèles ont dévié par rapport à la fréquence prévue (0,5). Ces déviations sont probablement dues à la dérive aléatoire. La méthode de comptage donne des résultats qui sont biaisés par excès. Le modèle de maximum de vraisemblance surévalue de 63% alors que le programme CRIMAP double la proportion d'haplotypes de recombinaison obtenue à partir de la base de données du génome de souris (MGD). Les résultats du CRIMAP suggèrent un ordre des marqueurs autre que celui mentionné dans MGD. Les erreurs de classification sont importantes à considérer dans l'analyse de telles données quoique la précision des estimations soit faible. Les résultats de cette étude mettent en évidence clairement une accumulation forte des haplotypes de recombinaison dans la région chromosomique étudiée.

Le **Chapitre 3** fournit des méthodes pour calculer la moyenne et la variance du nombre d'animaux ayant le génotype désiré à chaque génération de croisement de retour (backcross) dans une expérience d'introgression au moyen de marqueurs. Le but final est de produire des animaux homozygotes aux locus désirés. Les méthodes ont été développées spécifiquement pour des expériences avec les souches pures. Le modèle assume une distribution de Poisson pour la taille de la portée, et est semblable à cela utilisé dans les versions aléatoires des modèles de dynamique de population. Certains paramètres biologiques doivent être indiqués aussi bien que les paramètres sous le contrôle du sélectionneur. Ces méthodes peuvent être utilisées dans la phase de conception d'un programme pour déterminer le nombre d'animaux fondateurs, étant donné le nombre d'animaux voulus à l'accomplissement du processus de croisement et vice versa. L'attention est également portée sur la réduction au minimum du nombre total d'animaux à génotyper durant l'expérience, en variant le nombre de fois qu'un mâle sélectionné est utilisé. En outre, une méthode adaptative est proposée et qui tient compte des variations dans l'utilisation des mâles pendant l'expérience.

Pour optimaliser les modèles de programme d'introgression assistée par marqueur visant à introduire trois QTL non liés, le **Chapitre 4** a étudié des alternatives contre les phases de backcross ou d'intercross traditionnels. Les stratégies alternatives de backcross

sont plus avantageuses en ayant 50% moins de génotypage que la stratégie traditionnelle de backcross qui consiste à ne sélectionner que les animaux portant à la fois les trois QTL. Une phase de multiplication faisant suite à la sélection d'animaux homozygotes aux trois QTL lors du premier intercross permet de doubler leur nombre dans un modèle de souris. Dans ce même modèle, une deuxième génération d'intercross utilisant tous les animaux issus du premier intercross et ayant les trois QTL, aboutit à 12 fois le nombre d'animaux homozygotes obtenus à la première génération d'intercross. La diminution du nombre d'animaux à génotyper et du nombre de génotypage se situent dans les mêmes proportions quand le nombre d'animaux ciblés est fixe. L'une des options utilisant deux lignées d'animaux, chacune portant deux des trois QTL au cours de la phase de backcross et couplée avec la deuxième génération d'intercross, semble être la meilleure alternative d'introgression. Cette option nécessite 76% de moins pour le génotypage, 68% de moins pour le nombre d'animaux à génotyper et coûte 75% de moins qu'une option dans laquelle chaque animal est sélectionné sur la base des trois QTL.

Une expérience d'introgression de QTL de trypanotolérance au moyen de marqueurs génétiques a été conduite sur des souris de laboratoire dont la donneuse est la souche C57BL/6, et la receveuse la souche A/J. L'objectif était d'évaluer l'efficacité d'une telle introgression faisant usage des marqueurs génétiques et d'évaluer l'effet de chacun des trois QTL dans le nouveau génotype de fond. La stratégie de backcross qui consiste à sélectionner deux lignées de souris, chacune portant deux des trois allèles du QTL du donneur, a été utilisée. A la fin de la quatrième génération de backcross, des souris portant uniquement un allèle d'un QTL provenant du donneur ont été sélectionnées. A la fin de la phase d'intercross et selon le génotype des souris sur les chromosomes 1, 5 et 17, dix groupes de souris ont été constitués, plus les deux lignées de souris fondatrices comme des groupes de contrôle. Ces 12 groupes de souris ont été infectés de trypanosomes pour évaluer leur résistance. La méthode de génotypage successif a été appliquée à la troisième génération de la phase de backcross. Les résultats sur trois mois d'infection parasitaire ont été consignés dans le **Chapitre 5** de cette thèse. Ces résultats ont montré que les marqueurs génétiques sont transmis d'une génération à l'autre pendant la reproduction. Des goulots d'étranglement ont été observés, il s'agit de l'incapacité des animaux sélectionnés à survivre ou à se reproduire, de la disponibilité d'un nombre faible d'animaux ayant le génotype désiré, ou de la disponibilité d'animaux d'un même sexe (surtout crucial à la phase d'intercross). La méthode de génotypage successif réduit de plus de moitié le travail de génotypage. Les femelles ont une durée de survie supérieure à celle des mâles ($P < 0,001$). Des effets significatifs ont été observés pour chacun des trois QTL. L'effet génétique du QTL mesuré en nombre de jours de survie était de 10,0 sur le chromosome 1,

7,0 sur le chromosome 5, et 6,0 sur le chromosome 17. La proportion du génotype de fond de la souris donneuse a un effet très marqué sur la durée de survie à l'infection parasitaire. L'allèle du QTL provenant de la souris donneuse semble se comporter comme un allèle récessif sur les chromosomes 1 et 17 tandis qu'il est additif sur le chromosome 5. Les résultats de l'expérience décrite ici démontrent que l'introgression est faisable et pertinent dans un temps relativement court du moins avec les souris de laboratoire.

Au niveau du **Chapitre 6**, nous avons discuté de la contribution des chapitres deux à cinq à l'introgression chez les espèces animales en utilisant le backcross couplé avec les informations de marqueurs génétiques. Ensuite nous avons exploré les possibilités d'une introgression chez les bovins tenant compte des différentes méthodes de reproduction. Ultérieurement, nous avons passé en revue l'environnement sub-Saharien de l'élevage en Afrique avec insistance sur le système de reproduction. Le chapitre est clôturé avec quelques perspectives pour l'utilisation des gènes de trypanotolérance en Afrique au sud du Sahara. Une façon d'utiliser ces gènes est de faire la sélection pour des caractères de production à l'intérieur des animaux de race trypanotolérante. Dans le cas particulier de N'Dama, il serrait bénéfique de sélectionner pour des caractères de production de viande et de lait. Les gènes de trypanotolérance peuvent également être utilisés en exploitant des animaux hybrides de première génération obtenus en croisant les cheptels laitiers européens avec le bétail trypanotolérant local. Ce scénario est déjà en vigueur dans un grand nombre de pays en Afrique. L'introgression des gènes de trypanotolérance peuvent également être une bonne alternative pour améliorer la résistance aux trypanosomes des races exotiques de haut rendement importées en Afrique. Une autre alternative de l'utilisation des gènes de trypanotolérance est l'introgression de quelques gènes de production de lait dans les races reconnues trypanotolérantes en vue d'améliorer leur production. Etant donné la situation économique des pays sous l'emprise de la trypanosomose, une assistance financière internationale s'avère nécessaire pour entreprendre un programme d'introgression des gènes de trypanotolérance. Un tel programme ne devrait pas être une affaire d'un pays individuel, mais plutôt une responsabilité régionale.

Samenvatting

Trypanosomosis is een belangrijke en in Afrika wijdverspreide ziekte die zowel bij mensen als bij dieren voorkomt. Het heeft enorme gevolgen voor de landbouw vanwege het effect op mensen en dieren. De ziekte heeft bij de mens een afname in arbeidsproductiviteit tot gevolg en voorkomt dat land gebruikt kan worden voor de landbouw in gebieden waar de tsetse-vlieg veel voorkomt. Bij landbouwhuisdieren heeft trypanosomosis een negatief effect op de productie en kan fataal zijn als ze niet met medicijnen behandeld worden. Er worden geen nieuwe medicijnen ontwikkeld en de trypanosomen worden in toenemende mate resistent tegen de medicijnen die nu beschikbaar zijn. Er is nog geen vaccin beschikbaar en pogingen om de tsetse-vlieg uit te roeien is niet succesvol gebleken. De ziekte kost de Afrikaanse veehouders en consumenten jaarlijks ongeveer 1340 miljoen US dollars. Dit geeft aan dat de goed aangepaste runderrassen (N'Dama, Baoulé, Lagune, Muturu) veelbelovende genetische bronnen zijn voor de veehouderij.

Landbouwexperts hebben aangegeven dat vanwege trypanosomosis, de vochtige en semi-vochtige zones in Afrika, welke overeenkomen met de zones waar de tsetse-vlieg veel voorkomt, 18 tot 37% van de huidige productiecapaciteit niet wordt benut. Wanneer trypanosomosis kan worden gecontroleerd zouden naar schatting ongeveer 7 miljoen vierkante kilometer kunnen worden gebruikt voor de veehouderij of voor de gemengde bedrijfsvoering zonder dat dit veel nadelige effecten zou hebben voor het milieu. De introductie van een aantal lokale of synthetische rassen die trypanotolerant zijn zou in dat geval de opbrengst van de veehouderij kunnen verhogen.

Een duurzame optie om trypanosomosis te controleren is het houden van trypanotolerant vee. Experimenten om genen in kaart te brengen die gerelateerd zijn aan trypanotolerantie in N'Dama runderen met behulp van koppelingsanalyse zijn al in een vergevorderd stadium. Identificatie van de trypanotolerantie genen maakt de weg vrij voor merker-ondersteunde-introgressie. Het algemene doel van dit proefschrift is om optimale fokprogramma's te ontwikkelen voor de introgressie van trypanotolerantie genen, met gebruikmaking van de muis als model dier.

Voor introgressie van genen is het noodzakelijk deze te identificeren en te lokaliseren op het genoom. Het gebruik van meerdere generaties van AILs (Advanced Intercross Line) heeft tot doel het aantal recombinaties te verhogen, waarmee het mogelijk wordt de genen nauwkeurig in kaart te brengen. In hoofdstuk 2 wordt ingegaan op het aandeel recombinante haplotypes in een intercross F_9 muizenpopulatie (A/J x C57BL/6). Analyses zijn gebaseerd op genotyperingen in de F_9. Het deel van het genoom wat

onderzocht is, is ongeveer 7cM lang, ligt op chromosoom 17 en bevat een groot QTL gerelateerd aan trypanotolerantie in muizen. Het geschatte aandeel recombinante haplotypes is verkregen door simpelweg te tellen en door gebruik te maken van een maximum likelihood model dat geen rekening houd met fout typeringen. Om de eigenschappen van beide methodes te evalueren is een simulatiestudie uitgevoerd. Daarnaast is het CRIMAP programma gebruikt om deze data te analyseren. Allel-frequenties in de F9 wijken af van de verwachtte allel-frequentie van 0.5. Deze afwijkingen worden waarschijnlijk veroorzaakt door random drift. De tel methode geeft schattingen voor het aandeel recombinante haplotypes in de F_9 die systematisch te hoog zijn. Resultaten van het maximum likelihood model suggereren 63% meer recombinante haplotypes, terwijl de CRIMAP analyse recombinatie fracties geeft die twee keer zo hoog zij als de waardes die staan vermeld in de literatuur. De huidige dataset suggereert een merker volgorde die afwijkt van de literatuurgegevens. Misclassificatie is een belangrijke factor bij het analyseren van zulke datasets, ook al is de nauwkeurigheid van de schattingen van deze fractie laag. De resultaten van deze studie laten zien dat er duidelijk bewijs is dat in de F_9 het aantal recombinante haplotypes in het beoogde chromosomale gebied is toegenomen.

In hoofdstuk 3 zijn methodieken ontwikkeld om in elke willekeurige generatie van een merker-ondersteund terugkruisingsexperiment het gemiddelde en de variantie te berekenen van het aantal dieren met het gewenste genotype. Het uiteindelijke doel is dieren te produceren die homozygoot zijn voor de gewenste loci. De methodes zijn speciaal voor experimenten met ingeteelde lijnen ontworpen. Het model neemt een Poisson verdeling aan voor de worpgrootte en is vergelijkbaar met stochastische populatie dynamica modellen. Bepaalde biologische parameters en parameters die onder controle van de fokker staan moeten worden gespecificeerd. De ontwikkelde methodes kunnen worden gebruikt bij het ontwerpen van een merker ondersteund introgressie experiment. Een belangrijke vraag daarbij is het aantal dieren dat nodig is om een introgressie experiment op te starten gegeven het aantal dieren met het gewenste genoptype aan het eind van het experiment, en omgekeerd. Er is ook gekeken naar het minimaliseren van het totaal aantal genotyperingen gedurende het gehele experiment, door het aantal keren dat elk backcross mannetje is gebruikt te variëren. Tevens wordt er een beschrijving gegeven van een flexibel ontwerp, welke het mogelijk maakt het gebruik van mannetjes aan te passen gedurende het experiment.

Om het ontwerp van programma's voor merker-ondersteunde introgressie van drie ongekoppelde genen (QTLs) te optimaliseren, is in hoofdstuk 4 een vergelijking

gemaakt van verschillende alternatieven met een traditionele strategie. De alternatieve terugkruisingsstrategieën blijken efficiënter te zijn omdat ze 50% minder genotyperingen vragen dan een traditioneel terugkruisingsschema waarbij drie genen binnen één lijn worden gevolgd. Vergeleken met de eerste intercross generatie resulteert in een muis model het alternatief met een vermenigvuldigingsfase, volgend op de selectie van dieren die homozygoot zijn voor de drie genen, in een verdubbeling van het aantal homozygoten. Binnen hetzelfde model resulteert een tweede alternatief, waarbij de dieren die voor alle drie genen de gunstige allelen dragen, in de eerste intercross, in een 12-voudige toename in het aantal homozygote dieren dat wordt verkregen uit de eerste intercross generatie. Een afname in dezelfde orde van grootte in het aantal dieren dat gegenotypeerd moet worden en in het aantal genotyperingen dat uitgevoerd moet worden kan worden waargenomen, wanneer een vast aantal homozygote dieren wordt nagestreefd. De optie met twee lijnen die elk twee genen dragen gedurende de backcross generaties en gekoppeld zijn aan het tweede intercross alternatief, blijkt het beste introgressie alternatief te zijn. Deze optie vraagt 76% minder genotyperingen, 68% minder dieren die gegenotypeerd moeten worden en 75% lagere kosten in vergelijking met een optie waarbij alle drie de QTL worden ge-introgressed via een enkele lijn.

Een merker ondersteund introgressie (MAI) experiment is uitgevoerd om trypanotolerantie allelen van muis donor lijn, C57BL/6, in een ontvanger muis lijn, A/J over te brengen. Het doel was om een idee te krijgen over de effectiviteit van introgressie met behulp van genetische merkers en om de expressie van elk van de drie genen in hun nieuwe genetische achtergrond te testen. We hebben een terugkruisingsstrategie gebruikt die bestond uit het selecteren van twee lijnen, die elk twee van de donor allelen droegen gedurende de terugkruisingsfase (BC). In de vierde generatie werden dieren die voor één gen het gunstige allel droegen geselecteerd voor de intercross fase.
Gegeven het genotype op chromosoom 1, 5 en 17 zijn aan het eind van de intercross fase tien groepen muizen en twee controle groepen (C57BJ/6 en A/J) gevormd en vervolgens besmet met *Trypanosoma congolense*. Stapsgewijze genotypering is toegepast op de dieren in de BC3. In hoofdstuk 5 worden de resultaten gegeven die verkregen zijn gedurende de eerste 90 dagen post-inoculatie. De resultaten laten zien dat merkers overerven van de ene generatie op de volgende en gevolgd konden worden gedurende de fokkerij fase. Tijdens deze fase werden een aantal problemen geconstateerd: geselecteerde dieren die niet overleefden of niet tot reproductie in staat waren, de beschikbaarheid van maar weinig dieren van het gewenste genotype en beschikbaarheid van dieren van hetzelfde geslacht, wat cruciaal is gedurende de intercross fase. Stapsgewijze genotypering reduceerde het aantal noodzakelijk genotyperingen met meer dan 50%. Vrouwtjes hadden

een langere gemiddelde overlevingstijd dan mannetjes ($p<0.001$). Significante effecten zijn gevonden voor alle drie de genen. Het additieve effect van de genen op de overleving gemeten in dagen was 10.0 voor het gen op chromosoom 1, 7.0 voor het gen op chromosoom 5 en 6.0 voor het gen op chromosoom 17. De mate waarin er nog achtergrond genen van de donor lijn aanwezig waren bleek een opmerkelijk effect op de overlevingstijd te hebben. De allelen van de donor op chromosoom 1 en 17 lijken een recessieve werking te hebben, terwijl het gen op chromosoom 5 een additief effect laat zien. Het huidige experiment laat zien dat introgressie bij laboratorium dieren effectief is.

In hoofdstuk 6 bediscussiëren we de bijdrage van de hoofdstukken 2 tot 5 aan de introgressie bij dieren met gebruikmaking van een backcross fokprogramma en genetische merker informatie. Daarna hebben we aangegeven hoe introgressie gebruikt zou kunnen worden bij rundvee, gebruik makend van de huidige reproductie technieken. Vervolgens hebben we de veehouderij omstandigheden in landen in het sub-Sahara gebied bekeken waarbij we met name hebben gelet op het fokkerij systeem. Het hoofdstuk wordt afgesloten met een aantal vooruitzichten om trypanotolerantie genen in te benutten in het sub-Sahara gebied. Een mogelijkheid om trypanotolerantie genen te gebruiken is de selectie op productie kenmerken binnen de trypanotolerante rassen. In het speciale geval van de N'Dama zou het voordelig kunnen zijn om te selecteren op melk en vlees kenmerken. Trypanotolerantie genen kunnen ook gebruikt worden bij de exploitatie van F_1 dieren, geproduceerd uit een kruising tussen Europees melkvee met lokale tryapnotolerante rassen. Dit scenario wordt al toegepast in een groot aantal landen in Afrika. Introgressie van trypanotolerantie genen zou ook een goed alternatief kunnen zijn om resistentie te verhogen in exotische, hoog productieve melkvee rassen, geïmporteerd in Afrika om melkproductie te verhogen. Om een introgressie programma te starten zou internationale financiële hulp welkom zijn, gegeven de economische situatie in de verschillende betrokken landen. Een fokprogramma voor trypanotolerante runderen om melk en vlees te produceren zou niet alleen zaak moeten zijn voor het land zelf, maar een regionale verantwoordelijkheid.

Curriculum vitae

Olorounto Delphin Koudandé was born on July 18, 1955 in Toffo, Republic of Bénin. After completing the conventional grammar school in 1975, he started studies at the Faculty of Agriculture of the National University of Bénin. He completed his compulsory military service during the academic year 1977-1978. In 1978 he enrolled at the veterinary school in Dakar-Sénégal where he obtained his diploma of Doctor in Veterinary Medicine in June 1983. He went back to his country where he worked successively as researcher (1983-1987), director of a state-owned farm (1987-1988), head of animal health division in a regional extension service (1988-1989), and head of animal health division in the National Institute of Agricultural Research (1989-1995). In August 1995 he joined Wageningen University where he obtained a MSc degree with distinction in Animal Sciences with a major in epidemiology, in 1997. Meanwhile he started his PhD program in the Animal Breeding and Genetics Group of WIAS under the supervision of. Profs. Johan A.M. van Arendonk and Pim Brascamp. Delphin is returning to his position at the National Institute of Agricultural Research in Bénin after graduation.

Contact in Bénin

Office: INRAB (Institut National des Recherches Agricoles du Bénin)
01 BP 884 COTONOU
République du Bénin
Tél. 00 229 30 02 64 (deliver message)
Fax. 00 229 30 07 36

Home: 08 BP 268 COTONOU
République du Bénin
Tél. 00 229 36 01 89

e-mail: dkoud@hotmail.com