THE GENETIC ARCHITECTURE OF THE DDK SYNDROME:
AN EARLY EMBRYONIC LETHAL PHENOTYPE IN THE MOUSE

Folami Yetunde Ideraabdullah

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum in Genetics and Molecular Biology.

Chapel Hill
2007

Approved by:

Fernando Pardo-Manuel de Villena, PhD

Scott Bultman, PhD

Beverly Koller, PhD

Karen Mohlke, PhD

David Threadgill, PhD

ABSTRACT

Folami Yetunde Ideraabdullah: The Genetic Architecture of the DDK Syndrome: An Early
Embryonic Lethal Phenotype in the Mouse
(Under the direction of Fernando Pardo-Manuel de Villena)

The DDK syndrome is a polar early embryonic lethal phenotype that occurs when

DDK females are mated to males of other inbred mouse strains. Lethality is parent of origin

dependent and results from an incompatibility between an ooplasmic DDK factor and a non-

DDK paternal gene, both of which map to the *Ovum mutant* (*Om*) locus on chromosome 11.

Here, I utilize naturally occurring genetic variation in classical and wild-derived inbred strains

to characterize the genetic architecture of the DDK syndrome. I show that genetic variation

among wild-derived strains is uniformly distributed and significantly higher than previously

reported for other mammalian species. The high levels of diversity present among laboratory

strains suggests that the effective population size of the *Mus* lineage has been relatively

large and constant over a long period of time. Overall, these findings demonstrate that wild-

derived inbred strains are a valuable resource for genetic studies. By utilizing this resource

in recombination mapping and association mapping experiments, we have reduced the

candidate interval for the paternal gene of the DDK syndrome to a 23 kb region

encompassing a single gene. We have also defined a candidate interval for the gene

encoding the maternal factor, and demonstrated that the maternal and paternal components

of the DDK syndrome are non-allelic. I have identified three *Mus musculus domesticus* wild-

derived strains carrying modifiers that completely rescue the DDK syndrome lethality. In at

least two of these strains, the major modifier loci are unlinked to *Om* and rescue lethality in a parent of origin dependent manner that is independent of allelic exclusion at *Om*. Taken together, these data reveal that the DDK syndrome requires a specific combination of alleles at multiple loci. The fact that all of these alleles, with the exception of the allele encoding the maternal DDK factor, segregate in natural populations of mice suggests that they may be part of an important molecular pathway. In conclusion, further characterization of the genes responsible for this rescue phenotype will not only provide significant insight into the DDK syndrome, it should also increase our understanding of the molecular framework underlying early mammalian development.

DEDICATION

To my parents for overcoming the odds and living their lives as an inspiration to others;

and to my sister and three brothers for making me tough and believing in me.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

bp	Base pair(s)
chr	Chromosome(s)
df	Degrees of freedom
dpc	Days post coitum
E	Embryonic day
Gmnn	Geminin
H_o	Null hypothesis
IBD	Identical by descent
IBS	Identical by state
kb	Kilo base pair(s)
LOD	Logarithm of the odds
Mb	Mega base pair(s)
M. m.	Mus Musculus
n.a.	Not applicable
n.d. or N.D.	Not determined
Om	Ovum mutant
Rmod	Rescue modifier of the DDK syndrome
SD or Std Dev	Standard deviation
SDP(s)	Strain distribution pattern(s)
SEM	Standard error of the mean
Slfn	Schlafen
SNP(s)	Single nucleotide polymorphism(s)
TRD	Transmission ratio distortion

QTL(s)   Quantitative trait locus/loci

UTR   Untranslated region

CHAPTER 1. Background and Introduction

I. Specific aims

        During the course of my doctoral studies, I have completed the following specific aims:

        Aim 1. To characterize the level, spatial distribution, and phylogenetic history of genetic variation among classical and wild-derived inbred mouse strains using genomic sequence.

        Aim 2. To utilize naturally occurring genetic variation among inbred strains to reduce the candidate interval for the paternal gene responsible for the DDK syndrome through association and recombination mapping.

        Aim 3. To screen a diverse panel of F_1 hybrid females in order to identify inbred strains that carry dominant modifier alleles that rescue the DDK syndrome, and to characterize the extent, timing, and mode of rescue.

        Aim 4. To use a whole-genome scan to map major loci responsible for the rescue phenotype followed by high-resolution mapping to define candidate intervals; and also to determine whether modifiers present in the C57BL/6 strain are linked to the rescue modifier loci identified here.

        Aim 5. To determine whether allelic exclusion at *Om* is necessary for rescue and whether rescue is parent of origin dependent.

Aim 6. To identify candidate genes within the defined candidate interval for the major locus responsible for the rescue-of-lethality phenotype based on their expression profile, known or predicted interactions with genes at *Om*, and/or a known or predicted role in embryonic development.

Ultimately, the aim of the research presented here is to use the DDK syndrome as a model to gain insight into the essential genetic and epigenetic processes required for normal mammalian early embryonic development.

II. Early embryonic mouse development

Early embryonic development in mammals is an intricate process that involves a complex network of genetic and epigenetic interactions, many of which are timing and cell specific [1]. Although numerous studies have provided much insight into later development, we still do not fully understand the early developmental interactions that enable a single cell to differentiate into a fully functioning, multi-cellular organism. The events occurring prior to implantation are essential for normal development and play a major role in determining the expression profile, as well as the morphology and organization of the embryo later in development [1-3]. The "quality" of the preimplantation embryo is thought to determine its success at later stages of development and defects in the embryo occurring prior to implantation may contribute to the high levels of implantation failure in humans [4-7]. Such defects are also being implicated in an increasing number of postnatal abnormalities [8-12]. Therefore, further understanding of the molecular framework underlying preimplantation development is relevant in the etiology of human disease.

a) Mammalian gametes

Mammalian reproduction requires the fusion of male and female gametes, the sperm and the egg, respectively. Mammalian sperm consists of little more than a haploid nucleus

coupled with a few structures and enzymes that the sperm uses to locate and enter the egg. On the other hand, the egg, which is enveloped by the protective zona pellucida (Figure 1.1), contains a substantial cytoplasm consisting of most of the materials necessary for early embryonic development. These include proteins, ribosomes, tRNA, mRNA, as well as morphogenetic factors necessary for cell differentiation [13-15].
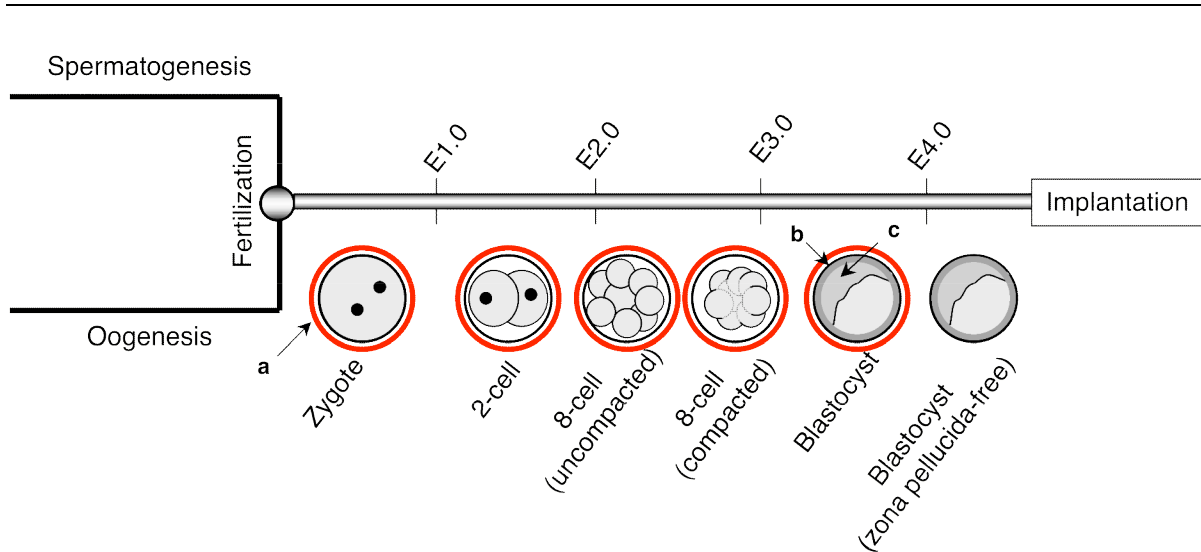
FIGURE 1.1. Preimplantation development of mouse embryos. (a) Red circles represent the zona pellucida; (b) trophectoderm; (c) inner cell mass.

b) Fertilization

Fertilization is the fusion of the sperm and egg, which leads to embryogenesis. It is made possible when enzymes present in the acrosome of the sperm digest the zona pellucida allowing the sperm to successfully penetrate the zona pellucida and fuse to the plasma membrane of the oocyte. This triggers a chain of events that mark the start of development beginning with egg activation. It has been proposed that upon binding to the oocyte, the sperm introduces a novel form of phospholipase C to the oocyte, which initiates production of inositol 1,4,5 triphosphate (IP_3) at the sperm entry site. IP_3 production induces the increase in intracellular calcium needed for egg activation [16-20].

Egg activation is an essential step during development and is necessary for the completion of meiosis II of the oocyte, block to polyspermy, and initiation of egg metabolism. Prior to fertilization, the oocyte is arrested in metaphase of the second meiotic division. The interaction between the sperm and oocyte triggers the completion of meiosis II, causing the oocyte to extrude the second polar body. Furthermore, the increase in intracellular calcium causes the release of cortical granule enzymes, which modify the proteins in the zona pellucida that bind to sperm receptors, thereby, blocking the binding of additional sperm [21, 22]. Soon after these events, activation of amino acid transport and protein synthesis occurs, as well as initiation of DNA synthesis in preparation for the first mitotic division [23].

Although the entire sperm enters the oocyte, only the genomic DNA and a few other components of the sperm head are incorporated into the egg. The other sperm components, including the mitochondrial DNA, are degraded. Prior to the first mitotic division, the maternal and paternal haploid complements are contained in separate pronuclei. The typical diploid nucleus is not formed until after the first cell cleavage [24, 25].

c) Preimplantation development

During the first five to six days of embryonic development in the mouse, the embryo goes through several essential stages, which are initially defined by the number of cell divisions that have occurred, and later by the morphology of the embryo. The cells of the early embryo divide asynchronously and by means of rotational, holoblastic cleavage [26, 27], and growth of the preimplantation embryo is restricted by the presence of the zona pellucida. Therefore, despite multiple cell divisions, the embryo remains the same size as that of the fertilized egg, until it hatches from the zona pellucida. The first cell cleavage occurs approximately 24 hrs after fertilization (Figure 1.1). Up until the 2-cell stage, when the embryo's genome is activated, embryonic development depends solely on maternal stores of mRNA and proteins that were accumulated by active transcription in the oocyte and

surrounding granulosa cells during oogenesis. After fertilization, this pool of mRNA is targeted for extensive degradation. By the 2-cell stage, approximately half of the maternal mRNA has been degraded, and most of it has been degraded by the 4-cell stage [28, 29]. Not surprisingly, studies suggest that the maternal mRNA is selectively degraded based on the role it plays in development [30]. Hence, maternal mRNAs involved in oogenesis are degraded rapidly, while maternal mRNAs involved in preimplantation development are degraded at a slower rate and many are replenished after zygotic genome activation [30-32].

The 8-cell, or morula stage, is marked by the compaction event in which the embryo goes from a loose arrangement of cells to a compact ball of cells (Figure 1.1). This compact structure is stabilized by tight junctions that form between the outer cell surfaces [33, 34]. Gap junctions formed between the inner cell surfaces allow small molecules and ions to pass between the cells [35]. In the next round of cell divisions, which gives rise to 16 cells, gap junctions and tight junctions are thought to play an important role in maintaining the distinct identities of the outer cells *vs.* the inner cells. This distinction is thought to have significant contribution to the first cell differentiation events in the embryo that occur after compaction [34]. Following compaction, the embryo continues to undergo cleavage and the cells differentiate to form two distinct cell types, the trophoblast cells of the trophectoderm, and the inner cell mass. While the trophectoderm gives rise to extraembryonic tissues, the inner cell mass gives rise to the embryo proper as well as some extraembryonic tissues (Figure 1.1). Proper differentiation and proliferation of trophoblast cells is necessary for successful implantation into the uterus.

The early blastocyst has a lining of trophoblast cells, an inner cell mass, and has begun to develop a fluid filled blastoceol cavity [36, 37]. By this time, the embryo has traveled from the oviduct, where fertilization occurred, to the uterus. Upon arrival in the

5

uterus, the blastocyst must be able to hatch from the zona pellucida and implant into the uterine wall for survival and further development (Figure 1.1).

d) Genetic and epigenetic interactions in the preimplantation embryo

Genetic and epigenetic interactions in the preimplantation embryo play a major role in development. These include the mitotic events that take place in the zygote and lead to the first cleavage event. Genome-wide epigenetic reprogramming of the parentally established expression profile also occurs during preimplantation development [38, 39]. DNA methylation is a widely recognized epigenetic modification of DNA associated with gene expression. At fertilization, both sperm and egg genomes are highly methlyated [38, 39]. However, by the 2-cell stage, the paternal genome is demethylated and the maternal genome is demethylated by the blastocyst stage. Remethylation of maternal and paternal genomes occurs around the time of implantation at different levels in the embryonic *vs.* the extraembryonic cell lineages. Despite this "resetting" of parental epigenetic marks, the mechanism by which the embryo's maternal genome is distinguishable from the paternal genome, is preserved. This parent of origin specific imprint is established during gametogenesis and is maintained in a subset of autosomal genes and also in the X chromosome of females.

During preimplantation development, cells of the embryo also undergo X-inactivation, a dosage compensation mechanism in female mammals which involves silencing of the genes expressed on a single X chromosome per cell. The major gene responsible for X-inactivation, *Xist*, is expressed from the paternal X chromosome as early as the 2-cell stage [40, 41]. Initially, the paternally inherited X is inactivated in all cells. While this imprinted X-inactivation is maintained in the extraembryonic cell lineages, it is lost in the embryonic cell lineages, which are subject to stochastic X-inactivation, such that one of the two X chromosomes is inactivated per cell regardless of parental origin [42, 43].

Epigenetic reprogramming, imprinting and X inactivation are essential to normal development. The genome's failure to undergo proper epigenetic reprogramming is thought to be one of the main causes of poor efficiency in somatic cell nuclear transfer (SCNT) cloning experiments [44-46]. Defects involving imprinted genes are associated with several human diseases including Beckwith-Wiedemann, Prader-Willi and Angelman syndromes as well as some types of cancer [47, 48]. Finally, skewed X inactivation has been found to be associated with ovarian and breast cancer [49].

III. The DDK syndrome

The DDK syndrome is a polar, parent of origin dependent lethal phenotype that disrupts early embryonic mouse development [50]. This phenotype was first described by Tomita in 1960 as a reduction in fertility observed in crosses between females of the DDK inbred mouse strain and males of other inbred strains [51]. He and others showed that when DDK females are mated to non-DDK males, a significant reduction (90-100%) in litter size is observed and in many cases impregnated females never give birth to live young [51, 52]. On the other hand, both the reciprocal cross (non-DDK female x DDK male) and the intra-strain cross (DDK female x DDK male) are fully viable (Table 1.1) [51, 52]. Crosses between F_1 hybrid females and incompatible non-DDK males as well as crosses between DDK females and F_1 hybrid males are described as being semi-sterile (50% lethality); while F_1 intercrosses are described as being 25% lethal (Table 1.1) [53]. It is important to note that (DDK x nonDDK)F_1 males and females have the same reproductive performance as (non-DDK x DDK)F_1 males and females, respectively (*n.b.* for all crosses described in this report the female is always listed first and male second unless otherwise noted) (Table 1.1).

Cytoplasm and pronuclear transfer experiments (discussed below) have shown that the polarity (*i.e.* the uni-directionality) of the lethal phenotype is due to the fact that lethality

TABLE 1.1. Reproductive performance of crosses involving C57BL/6(B6) and DDK(K).

Female	x	Male	% embryo survival	Expected genotypes at *Om* of surviving embryos*	References[‡]
B6		B6	100%	100% B6/B6	[54]
DDK		DDK	100%	100% K/K	[51, 54]
B6		DDK	100%	100% B6/K	[54]
DDK		B6	0-10%	100% K/B6	[54]
(B6 x DDK)F$_1$		B6	50%	50% K/B6, 50% B6/B6	[53]
(B6 x DDK)F$_1$		DDK	100%	50% B6/K, 50% K/K	[53]
(DDK x B6)F$_1$		B6	50%	50% K/B6, 50% B6/B6	[55]
(DDK x B6)F$_1$		DDK	100%	50% B6/K, 50% K/K	[55]
B6		(B6 x DDK)F$_1$	100%	50% B6/K, 50% K/K	[53]
DDK		(B6 x DDK)F$_1$	50%	~100% K/K[†]	[53]
B6		(DDK x B6)F$_1$	100%	50% B6/K, 50% B6/B6	[55]
DDK		(DDK x B6)F$_1$	50%	~100% K/K[†]	[55]
(B6 x DDK)F$_1$		(B6 x DDK)F$_1$	75%	33.3% B6/K, 16.7% K/B6, 33.3% K/K, 16.7% B6/B6,	[56]

Maternally inherited allele is listed first followed by the paternally inherited allele.
*Expectations based on Wakasugi's genetic interpretation [53].
[†]A few embryos with incompatible genotypes (K/B6) survive.
[‡]Reference for cross description.

is the result of an incompatibility between an ooplasmic DDK factor and a non-DDK paternal

gene [57, 58]. Furthermore, these experiments demonstrated that the lethal effect is

dependent on the parental origin of the non-DDK gene, such that the incompatibility only

occurs between DDK ooplasm and a paternal non-DDK pronucleus but not between DDK

ooplasm and a maternal non-DDK pronucleus [57].

The parent of origin dependent, heterozygote-specific nature of the DDK syndrome

has been used as a reference to define the polar overdominance muscular hypertrophy

phenotype associated with the ovine *callipyge* locus [59]. This phenotype occurs as a result

of heterozygosity for wild-type and mutant alleles at the *callipyge* locus, but only when the

mutant allele is transmitted through the sire. No phenotype is observed when offspring are homozygous for the mutant allele, nor when homozygous for the wild-type allele. The *callipyge* phenotype has been proposed to be the result of a *trans* interaction between reciprocally imprinted genes, a maternally expressed repressor and its paternally expressed growth-promoting target [60].

a) Timing of lethality

It was determined that most (DDK x non-DDK)F_1 embryos exhibit developmental abnormalities as early as the 2-cell stage [61]. By counting the number of embryonic cells at 36 hrs, 60 hrs, and 84 hrs *post coitum*, researchers observed that (DDK x non-DDK)F_1 embryos progress to the 2-cell stage at the same rate as normal control embryos. However, in most cases, further cell cleavage in (DDK x non-DDK)F_1 embryos occurs at significantly slower rates and never recovers, resulting in significantly smaller cell numbers throughout the rest of early development. It is important to note that DDK embryos also have delayed transition from the 2-cell stage to the 4-cell stage, however, by 2.5 days post *coitum* (*dpc*) cell numbers are significantly higher than (DDK x non-DDK)F_1 embryos [61].

While the effects of the DDK syndrome are evident as early as the second cell cleavage event, most of the affected embryos arrest and begin to degenerate between the morula and peri-implantation stages [52, 54]. This was determined by comparing the number of *corpora lutea* present in the female to the number of morphologically "normal" embryos at different *dpc* after mating with incompatible non-DDK males. Females were dissected at different stages of pregnancy and normal *vs.* abnormal embryos were determined on the basis of expected cleavage, cell number, embryo morphology, or by staining with Nigrosin, to differentiate live *vs.* dead cells. By 3-4 *dpc* many of these embryos lacked a blastocoel, and had a poorly developed trophoblast and inner cell mass [54]. By 5

dpc, almost all of the embryos were classified as "disorganized" or "degenerate" having poorly developed primitive endoderm and inner cell mass. Although most (DDK x non-DDK)F$_1$ embryos do not progress beyond the egg-cylinder stage (~5 *dpc*), on rare occasion a few embryos survive to birth (Table 1.1). These offspring are fully fertile and do not develop any obvious postnatal defects.

b) Physiological cause of lethality

The direct physiological cause of embryonic death due to the DDK syndrome is unknown. However, previous experiments ruled out some possible causes such as defects in fertilization or unfavorable interactions between the embryo and the uterine environment. In crosses between DDK females and non-DDK males, normal numbers of eggs complete meiosis II and proceed with the first cell cleavage event [61]. Furthermore, ovary transplantation experiments demonstrated that the uterine environment is not the cause of embryonic lethality as (DDK x non-DDK)F$_1$ embryos died regardless of the uterine host [54].

Affected embryos exhibit a range of phenotypes. Morphologically, these embryos become progressively abnormal as the complexity of development increases. As discussed above, embryos affected by the DDK syndrome have significantly slower cleavage rates compared to normal embryos as well as defects in cell differentiation that lead to undeveloped trophoblast and inner cell mass. Also, few affected embryos hatch from the zona pellucida (Figure 1.2) [57]. Low intracellular pH leading to defective gap junction communication is a common feature of (DDK x non-DDK)F$_1$ embryos and has been proposed to be involved in the lethal phenotype [62, 63]. Gap junctions play an integral role in compaction of the 8-cell embryo [33]. Although (DDK x non-DDK)F$_1$ embryos have been shown to undergo compaction at the 8-cell stage, in the late 16-cell stage, for unknown reasons they decompact [62].

Due to the differences in timing of lethality, as well as the range of phenotypes observed to be associated with the DDK syndrome, it has been difficult to determine the primary physiological cause of death.

c) Genetic basis of lethality

Based on the polarity of the DDK syndrome, Tomita proposed that lethality is caused by an incompatibility between the egg cytoplasm of DDK females and "alien" or non-DDK spermatozoa [51]. This hypothesis was confirmed by pronuclear transplantation and ooplasm transfer experiments that showed that the incompatibility is not between the
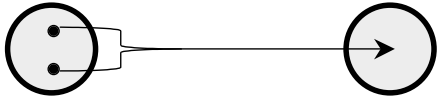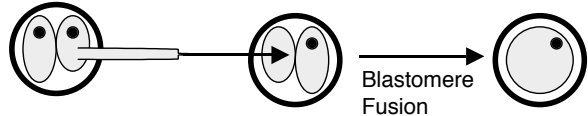
	Pronuclear Donor	Enucleated Recipient	Results	References
a.	(CBA x B6)F_1	(CBA x B6)F1	93% embryos at E10.5-15.5	[58]
b.	[(CBA x B6) x (CBA x B6)]F_2	[(CBA x B6) x (CBA x B6)]F_2	55% embryos at E10.5-15.5	[58]
c.	[DDK x (CBA x B6)]G_2	[(CBA x B6) x (CBA x B6)]F_2	70% embryos at E10.5-15.5	[58]
d.	(DDK x DDK)F_1	(DDK x DDK)F_1	82% blastocysts 64% hatched	[57]
e.	(DDK x BALB/c)F_1	(BALB/c x DDK)F_1	72% blastocysts 55% hatched	[57]
f.	(DDK x BALB/c)F_1	(DDK x DDK)F_1	16% blastocysts 0% hatched	[57]
g.	(BALB/c x DDK)F_1	(DDK x DDK)F_1	81% blastocysts 57% hatched	[57]
h.	(B6 x DDK)F_1	(DDK x BALB/c)F_1	48% blastocysts 14% hatched	[57]
i.	(B6 x DDK)F_1	(DDK x DDK)F_1	92% blastocysts 76% hatched	[57]
j.	(BALB/c x DDK)F_1	(DDK x BALB/c)F_1	36% blastocysts 18% hatched	[57]

FIGURE 1.2. Pronuclear transplantation experiments. For all crosses, female is listed first and male second. The first and second columns list crosses from which pronuclear donor embryos and enucleated recipient embryos, respectively, were harvested. The third column lists the percentage of embryos that developed into blastocysts, hatched from the zona pellucida, or survived past E10.5 out of the total number of transplanted samples.

pronuclei [58], but is between the DDK ooplasm and the non-DDK pronucleus (Figure 1.2 and Figure 1.3) [50, 57]. These studies also demonstrated that lethality is parent of origin dependent such that the lethal interaction is only observed between the DDK ooplasm and a paternal non-DDK pronucleus, but not between the DDK ooplasm and a maternal non-DDK pronucleus (Figure 1.2d-i; Figure 1.3a & b) [50, 57]. Furthermore, it was shown that the incompatible interaction occurs by the one-cell stage and is not reversible by simply transplanting compatible pronuclei into the affected 1-cell embryos (Figure 1.2j) [57]. RNA microinjection experiments demonstrated that the DDK maternal factor is present in the DDK oocyte as an RNA that has the ability to affect embryonic development even when introduced into the embryo as late as the 4-cell stage [64]. Recently, it has been shown that while the incompatible interaction between the ooplasm and the paternal genome is recapitulated in cloned embryos, the parent of origin effect is not; indicating that the

Cytoplasm Donor	Enucleated Recipient	Results	
a. (DDK x DDK)F_1	(BALB/c x BALB/c)F_1	37% blastocysts	Modified from [50]. The first and second columns list crosses from which 2-cell cytoplasm donor embryos and recipient 2-cell embryos having a single enucleated blastomere, respectively, were harvested. The third column lists the percentage of embryos that developed into blastocysts, out the total number of transferred samples.
b. (DDK x DDK)F_1	(BALB/c x DDK)F_1	74% blastocysts	

Blastomere Fusion

epigenetic mark discriminating between maternal and paternal pronuclei is not retained in somatic cells [65]. Crosses between reciprocal F_1 females and incompatible males

12

demonstrated that the incompatible ooplasmic factor is not of mitochondrial origin as (non-DDK x DDK)F_1 females have the same reproductive performance as (DDK x non-DDK)F_1 females (Table 1.1) [50].

Because the lethal interaction involves both a maternal factor and a paternal gene, it has been deemed both a maternal and paternal effect [50]. The presence of a parent of origin effect suggests that differential expression of maternal and paternal genomes may play a role in lethality. A reverse imprinting mechanism was proposed to explain the parent of origin effect [66]. This hypothesis was later tested and rejected on the basis the segregation of alleles at *Om* in progeny testing (see below) [55, 67].

d) Mapping genetic components of the DDK syndrome

On the basis of progeny testing, the original genetic model of the DDK syndrome postulated that either a single gene or two tightly linked genes were responsible for the embryonic lethality [53]. Wakasugi named the gene encoding the maternal factor *ovum mutant* and used the abbreviations *om* and *OM* for the DDK and non-DDK alleles, respectively. Likewise, he named the alleles at the paternal gene *s* and *S* for the DDK and non-DDK allele, respectively. Under this model, the genotypes of the DDK strain at the gene encoding the maternal factor and the paternal gene would be (*om/om*, *s/s*), respectively; while incompatible strains would have the genotypes (*OM/OM*, *S/S*). Therefore, embryos carrying both the *om* and *S* alleles die (*i.e. om/OM*, *s/S*); however, embryos carrying either *om* or *S* alleles but not both (*i.e. om/OM*, *s/s or OM/OM*, *s/S*) survive. In later studies, investigators refer to the locus for both the paternal gene and the gene encoding the maternal factor as the *Ovum mutant* (*Om*) locus after a report by Baldacci and coworkers [68] states that their data supports the single locus hypothesis. However, direct evidence

that the two components are closely linked is not shown until a few years later [69]. The allelic nature of the maternal and paternal components was not addressed in this study.

The paternal gene

The paternal gene at *Om* was mapped to the distal portion of mouse chromosome 11 in two separate studies. In the first study [68], the reproductive performance of crosses between DDK females and ((BALB/c x (BALB/c x DDK))N$_2$ males was tested *in vivo* (average litter size) and *in vitro* (cultured embryos). Males were genotyped at markers across the genome and evidence for linkage was found on chromosome 11. Using recombinant inbred lines derived from BALB/c and DDK, it was confirmed that *Om* maps to a locus on chromosome (chr) 11 that is closely linked to *Sigje* (*Ccl2*).

In the second set of experiments [66], the paternal gene was mapped using backcrosses between DDK females and F$_1$ hybrid males. This was done under the expectation that embryos inheriting a paternal B6 allele at *Om* die, while embryos inheriting a paternal DDK allele survive (Table 1.1). Therefore, at loci closely linked to *Om*, there should be a significant excess of surviving embryos having inherited the paternal DDK allele *vs.* the paternal B6 allele. Hence, *Om* was mapped by testing for statistically significant departures from the expected Mendelian segregation ratios, or transmission ratio distortion (TRD) of the paternal DDK allele at loci across the genome. TRD was observed at three marker loci on chr 11, the most significant of which was 51 centimorgans (cM) from the centromere and had a transmission ratio of 26:152 for paternal B6 to DDK alleles. In conclusion, although one study tested for association between genotype and phenotype, and the other study tested for TRD, both studies mapped the *Om* locus to the distal portion of chr 11.

Progeny testing of BALB/c-DDK recombinant inbred lines and the construction of a YAC/BAC-based physical map of the *Om* region was later used to define a 1.5 Mb

14

candidate interval for the paternal gene located between *Scya2* and *D11Pas18* [70, 71].

This 1.5 Mb interval for *Om* contains 33 known transcripts, 5 novel transcripts, 1

pseudogene, 9 ESTs, and 22 novel Genescan predictions [72, 73].

Positional cloning of the paternal gene was hindered by the large number of genes

present in the candidate interval, the lack of obvious candidates, the early onset of the lethal

phenotype and the complex nature of the phenotype, requiring specific allelic contributions

in a parent of origin dependent manner.

The gene encoding the maternal DDK factor

Initial efforts to identify *Om* focused on mapping the paternal gene. It was assumed

the gene encoding the maternal factor and the paternal gene were either at the same locus

or closely linked based on Wakasugi's original hypothesis (discussed above). In 1997,

Pardo-Manuel de Villena and coworkers demonstrated that the gene encoding the maternal

factor was indeed closely linked to the paternal gene at *Om* [69]. In this study, the

reproductive performance of crosses between B6 males and Om^{B6}/Om^{B6} or Om^{B6}/Om^{DDK} N_2

females was determined on the basis of litter size distribution. The reproductive performance

of N_2 females was found to be correlated with their genotypes at *Om*. The litter size

distribution of Om^{B6}/Om^{B6} N_2 females was not significantly different from the fully fertile cross

between (B6 x DDK)F_1 females and DDK males. Likewise, the litter size distribution of

Om^{B6}/Om^{DDK} N_2 females was not significantly different from the semi-lethal cross between

(B6 x DDK)F_1 females and B6 males. These findings confirmed Wakasugi's hypothesis that

the loci responsible for the lethal phenotype are closely linked but did not determine whether

the gene encoding the maternal factor and the paternal gene are allelic or non-allelic [53].

Sapienza and coworkers [66] proposed that the parent of origin dependent nature of

the DDK syndrome could be explained by a reverse imprinting effect at a single gene (*i.e.*

Om; Table 1.2). In this model, the *Om* locus is subject to differential expression based on

the parental origin of alleles. In non-DDK strains a maternally inherited allele at *Om* is expressed while a paternally inherited allele at *Om* is silenced. It was proposed that this differential expression is reversed in the DDK strain such that the maternally inherited Om^{DDK} allele is silenced while the paternally inherited Om^{DDK} allele is expressed. Therefore, in crosses between (B6 x DDK)F$_1$ females and B6 males, the expectation is death of all Om^{DDK}/Om^{B6} embryos (because both alleles would be silenced); and survival of all Om^{B6}/Om^{B6} embryos (because the maternal B6 allele would be expressed) (Table 1.2). This hypothesis was rejected when later studies demonstrated that in crosses between

---

TABLE 1.2. Reverse imprinting hypothesis*

Epigenotype at *Om*				Genotypes at *Om* of surviving embryos		
Dam	Sire	Embryo	Phenotype	Expected	Observed	References†
B6 (+)	B6 (-)	+/-	100% viable	100% B6/B6	100% B6/B6	
DDK (-)	DDK (+)	-/+	100% viable	100% K/K	100% K/K	
B6 (+)	DDK (+)	+/+	100% viable	100% B6/K	100% B6/K	
DDK (-)	B6 (-)	-/-	90-100% lethal	100% K/B6‡	100% K/B6‡	
F$_1$ (+/-)	DDK (+)	+/+, -/+	100% viable	50% B6/K, 50% K/K	50% B6/K, 50% K/K	[66]
F$_1$ (+/-)	B6 (-)	+/-, -/-	50% lethal	~100% B6/B6‡	**40% B6/B6 60% DDK/B6**	[56]
B6 (+)	F$_1$ (-/+)	+/-, +/+	100% viable	50% B6/B6, 50% B6/K	n.d.	
DDK (-)	F$_1$ (-/+)	-/+, -/-	50% lethal	~100% K/K‡	~100% K/K‡	[66]
F$_1$ (+/-)	F$_1$ (-/+)	+/-, -/+, +,+, -/-	25% lethal	33.3% B6/B6, 33.3% K/K‡ 33.3% B6/K,	**12.0% B6/B6,** 33.3% K/K, **54.7% B6/K & K/B6**	[56]

Maternally inherited allele is listed first followed by the paternally inherited allele. Bold genotypes highlight when observed genotype percentages are different from expected.
(+) *Om* expressed; (-) *Om* silenced. n.d., Not determined.
F$_1$, (B6 x DDK)F$_1$
*Modified from [66].
†References for observed genotypes.
‡A few embryos with incompatible genotypes (K/B6) survive.

(B6 x DDK)F$_1$ females and B6 males, Om^{B6}/Om^{B6} embryos do not have a higher rate of survival than Om^{DDK}/Om^{B6} embryos (Table 1.2) [55]. This finding was replicated in crosses involving (BALB/c x DDK)F$_1$ females [67].

In support of the non-allelic model, males from the MOM and CASP inbred strains have been shown to be fully compatible in crosses with DDK females [74]. Although the authors failed to recognize the significance of this finding, it demonstrated that alleles at the gene encoding the maternal factor and the paternal gene segregate independently of each other. DDK carries the incompatible allele at the gene encoding the maternal factor and a compatible allele at the paternal gene. Strains such as B6 and BALB/c carry a compatible allele at the gene encoding the maternal factor and an incompatible allele at the paternal gene. MOM and CASP carry compatible alleles at both the gene encoding the maternal factor and the paternal gene. This provides evidence that the gene encoding the maternal DDK factor and the paternal gene are non-allelic.

Identification of the gene encoding the maternal DDK factor has proven difficult. This is mostly due to the fact that the maternal contribution to the DDK syndrome acts as an incompletely penetrant maternal effect. In crosses between (B6 x DDK)F$_1$ females and B6 males, approximately half of the surviving embryos inherit a maternal Om^{B6} allele and half inherit a maternal Om^{DDK} allele [55] (Table 1.2). These results demonstrated that embryo survival is independent of the embryo's genotype at Om. Therefore, the phenotype of the embryo must depend on the dam's genotype, and hence it is a maternal effect. Furthermore, it was concluded that the maternal effect is incompletely penetrant because only half of the embryos are affected in these crosses. To account for this incomplete penetrance, it was proposed that the gene encoding the maternal DDK factor undergoes random allelic exclusion, such that only one of the two alleles present is expressed in a given oocyte (Figure 1.4). This would generate two classes of oocytes in (B6 x DDK)F$_1$ females, one class

17

that expresses the DDK allele and dies upon fertilization by B6 sperm, and another that

expresses the B6 allele and survives (Figure 1.4) [55]. Under this hypothesis, the gene

encoding the maternal factor would have to be expressed prior to the first meiotic division of

the oocyte; as expression at a later time would not result in the transmission ratio of alleles

observed in these crosses. This is in agreement with the results of the cytoplasmic and

nuclear transfer studies, which demonstrate that the maternal DDK factor is present in the

cytoplasm of the oocyte (Figures 1.2 and 1.3).
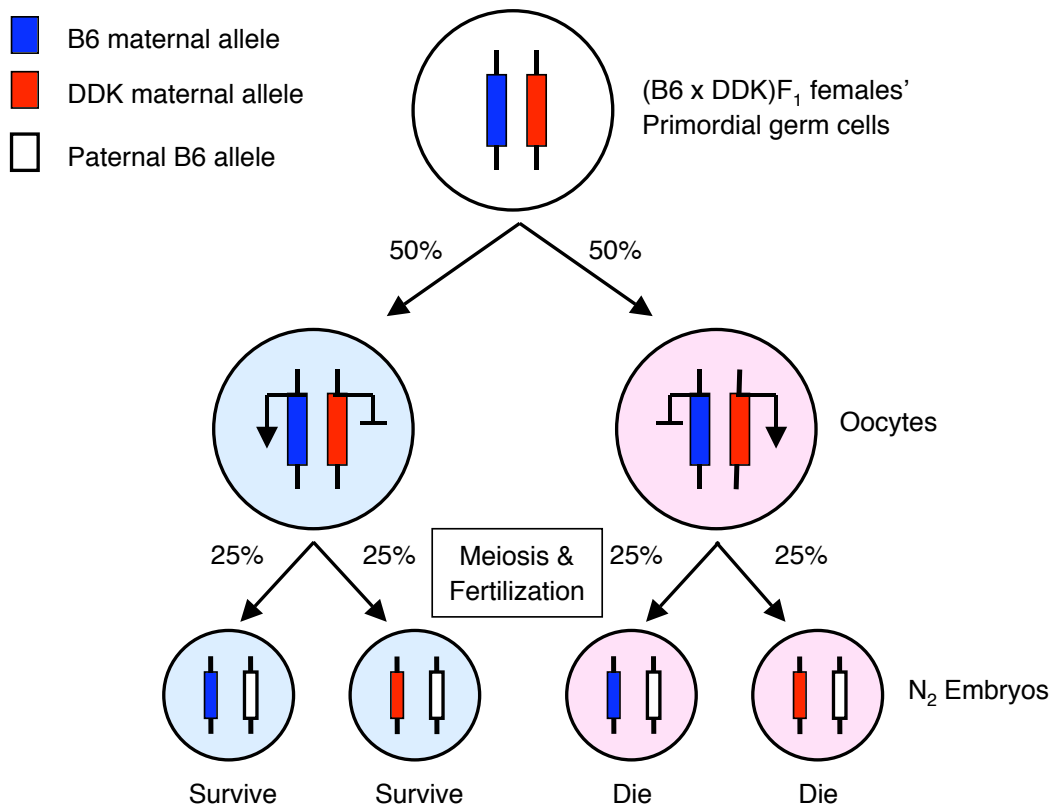
FIGURE 1.4. Allelic exclusion hypothesis.

18

e) Modifiers of the DDK syndrome

Pardo-Manuel de Villena and colleagues [75] showed that B6-DDK mixed background Om^{B6}/Om^{DDK} females mated to B6 males have wide variation in reproductive performance. The females generated by backcrossing to B6 had significantly lower reproductive performance when compared to (B6 x DDK)F$_1$ females or Om^{B6}/Om^{DDK} females generated by backcrossing to DDK. This correlation between the level of embryonic death and the proportion of the dam's genome that is of B6 or DDK origin was only observed in crosses involving females that were heterozygous at *Om*. It was attributed to the presence of multiple additive/epistatic, recessive, modifiers that are unlinked to *Om*. Additional studies replicated the findings of B6 modifiers [67, 76, 77]; and others demonstrated the presence of similar modifiers on the BALB/c strain background [77]. It was proposed that the presence of these modifiers increased (B6 and BALB/c) or decreased (DDK) the fraction of embryos that die due to the DDK syndrome. Mostly due to the fact that there are multiple additive/epistatic recessive modifiers, and that these modifiers cause an increase in lethality, the genomic locations of these modifiers have not been determined.

The presence of modifiers that increase or decrease lethality, in mixed background females that are heterozygous at *Om*, provides support for the allelic exclusion hypothesis discussed above. We and others have proposed that these modifiers act through allelic exclusion by skewing the choice of the allele to be expressed at the gene encoding the maternal factor and thereby causing fewer or greater numbers of oocytes to express the maternal DDK factor [67, 75, 77].

The molecular mechanism of allelic exclusion is still poorly understood, although, it is thought to be an important regulatory effect on expression of a growing number of genes in mammals [78-82]. Furthermore, aberrant or loss of allelic exclusion has been suggested to play a role in human disease [83-85].

f) Relevance of the DDK syndrome

The embryonic lethal DDK syndrome is caused by one of the earliest acting spontaneous mutations reported in an inbred mouse strain. It provides a unique angle with which to study early embryonic development in mammals and can be used to answer an array of fundamental biological questions pertaining to many aspects of early embryonic development such as cell cleavage, differentiation and fate, cell-cell communication, and implantation. Furthermore, the complex nature of this phenotype has been proposed to be due to the involvement of maternal effects and epigenetic effects such as allelic exclusion and imprinting. These aspects of the DDK syndrome are of particular interest due to their potential implications in human disease. Therefore, it is important to gain further understanding of the molecular and genetic basis of the DDK syndrome and how it leads to disruption of early embryonic development.

The DDK syndrome may also be useful in studying the mechanism of post-zygotic reproductive isolation under the Dobzhansky-Muller hypothesis [86, 87]. It could be speculated that the alleles at genes involved in the DDK syndrome might cause the type of hybrid incompatibilities that lead to speciation. Also, of equal interest is the meiotic drive phenotype that is closely linked to *Om*. Meiotic drive is defined as the non-random segregation of alleles between functional and nonfunctional products of female meiosis [88, 89]. This phenomenon results in TRD, which is in contrast with the expected 50:50 ratio, as expected under Mendel's laws of segregation. Meiotic drive has been shown to occur at *Om* such that upon fertilization of oocytes from (B6 x DDK)F_1 females by B6 sperm, the maternal *Om^{B6}* allele is preferentially segregated to the second polar body leaving the maternal *Om^{DDK}* allele in the zygote [55, 69, 90, 91]. This results in TRD in favor of the maternal DDK allele at ratios as distorted as 60:40. This form of TRD is not observed when oocytes from (B6 x

DDK)F$_1$ females are fertilized by DDK sperm. During the course of my doctoral studies, I contributed to a publication that showed that the alleles responsible for meiotic drive at *Om* occur at a high frequency in natural populations and segregate among *M. spicilegus* and *M. musculus* lineages [92]. The study of meiotic drive at *Om* yields insight into a unique mechanism of natural selection in which paternal contribution to the embryo influences the maternal genetic contribution.

IV. Laboratory inbred mouse strains as tools for genetic studies

Laboratory inbred mouse strains are recognized as highly valued tools for studying mammalian genetics. In general, mice are good research models due to the fact that it is relatively inexpensive to feed and house them, in part due to their small stature and short generation time. Furthermore, they are susceptible to genetic engineering and tolerate complete inbreeding, which among other things, allows for the stable maintenance of large collections of lines with spontaneous and induced mutations. Because they have many similar biological pathways and have similar genomes to humans, mice currently play a central role in studying a range of complex phenotypes related to human disease such as cancer and diabetes. This is greatly aided by the publicly available annotated sequence of the whole mouse genome.

More than 450 inbred mouse strains have been described [93]. They are commonly divided into two groups: "classical" and "wild-derived" inbred strains (Figure 1.5). Understanding the level, spatial distribution, and history of genetic diversity present among inbred mouse strains has important implications in evolutionary studies as well as in the experimental design of other genetic studies, such as those seeking to map complex traits.

Classical Inbred Mice

Wild-derived Inbred Mice

Wild mice (*Mus musculus*):
M. m. domesticus, *M. m castaneus*,
M. m. musculus, *M. m. molossinus*

Selection & Partial Inbreeding

European & Asian Fancy mice

Inbreeding

> 300 inbred strains

Wild mice:
M. musculus,
M. spicilegus,
M. spretus
M. caroli
M. pahari

Inbreeding
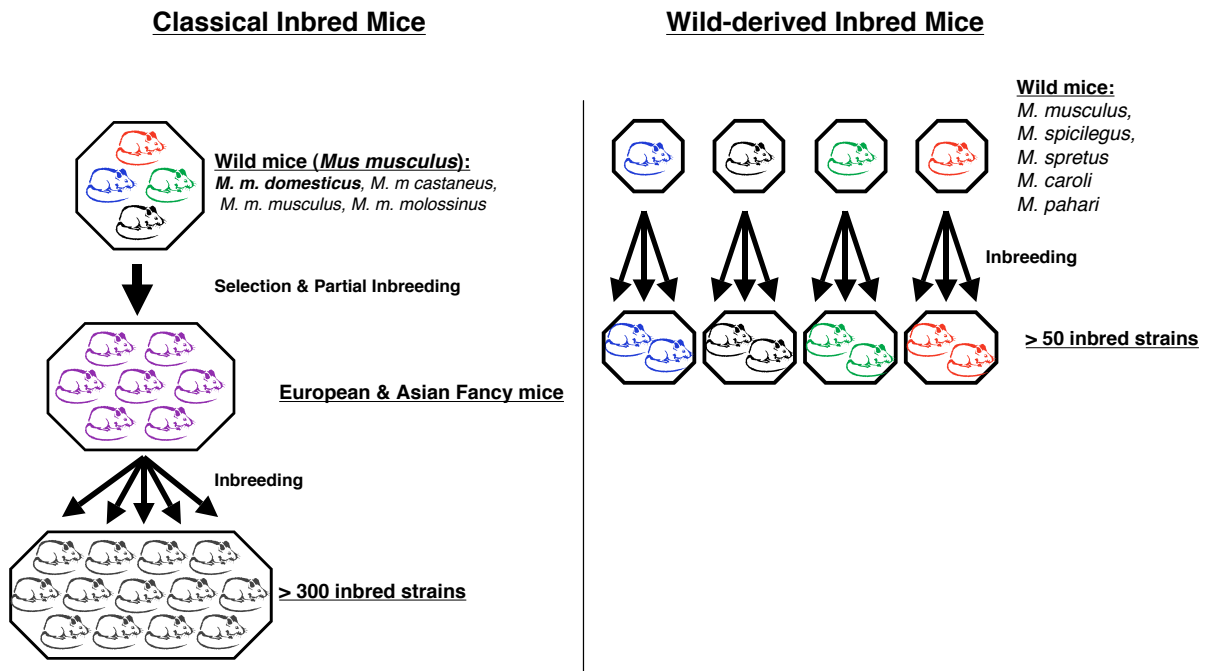
> 50 inbred strains

FIGURE 1.5. Derivation of classical and wild-derived inbred mouse strains.

a) Classical inbred strains

Classical inbred laboratory strains were derived about 100 years ago from "fancy" European or Asian mice that were bred for the purpose of being sold as pets (Figure 1.5). The genome of classical inbred strains derives from a handful of progenitors [93-95] and represents a mosaic with unequal contributions of the *Mus musculus* subspecies, *M. m. domesticus*, *M. m. musculus*, *M. m. castaneus*, and *M. m. molossinus* (Figure 1.5) [96]. The majority of the classical inbred strain genome is thought to be of *M. m. domesticus* origin [97, 98], although the mitochondrial genome and the Y chromosome of many classical strains were shown to be of *M. m. domesticus* and *M. m. musculus* origin, respectively [94, 99]. Recent reports indicate that classical inbred strains have limited levels of genetic variation when compared to humans and that the genetic variation is unevenly distributed,

22

with regions of low levels of polymorphism and regions of high levels of polymorphism [96, 97, 100-102]. Although the fraction of the genome that belongs to each category and the size of these regions remain under discussion, these observations have major implications for mapping quantitative trait loci [101-103].

DDK, C57BL/6, and BALB/c are all classified as classical inbred strains. The DDK inbred strain was derived along with several other strains in 1944 by K. Kondo at the University of Tokyo, Japan from a German mouse line called DD, for Deutsche Maus at Denken [93, 104-108]. Both C57BL/6 and BALB/c are very widely used in scientific research. They were derived at the Jackson Laboratory, Maine, USA in the early 1900s from mice obtained from the A. Lathrop pet mouse colonies in Granby, MA, USA [93, 104, 105].

b) Wild-derived inbred strains

Wild-derived inbred laboratory mice consist of several dozen inbred strains that have been derived from wild mice trapped at different times from different populations in distant geographical locations over the past 50 years (Figure 1.5) [93, 109]. These include inbred strains established individually from mice belonging to several *Mus* species, including *M. musculus*, *M. spretus*, *M. spicilegus*, *M. caroli* and *M. pahari*; as well as inbred strains derived from different *Mus musculus* subspecies and their inter-subspecific hybrids (Figure 1.5) [110]. The phylogenetic relationships and the divergence times among these species and subspecies have been established on the basis of DNA hybridization and limited sequence comparison studies (Figure 1.6) [109, 111-114].

Wild-derived inbred strains are thought to be particularly useful because they capture the genetic diversity of the individual species or subspecies of mouse from which they are derived. Therefore, high levels of genetic diversity are expected in interspecific and intersubspecific comparisons. Hybrids generated by crossing wild-derived and classical strains have been used to generate the high-resolution linkage map of the mouse, in the
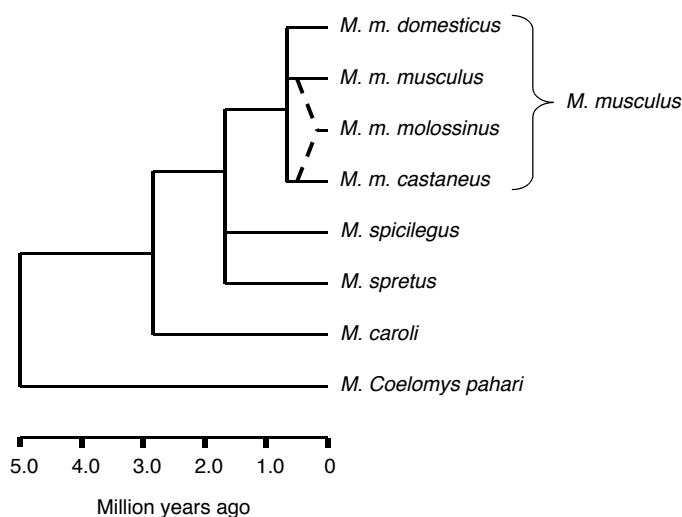
FIGURE 1.6. Phylogenetic tree of the *Mus* species and subspecies. The tree depicts a simplified phylogeny of *Mus* species and subspecies from which commonly used classical and wild-derived laboratory inbred mouse strains were derived. Modified from [109].

study of genomic imprinting, X-inactivation and complex traits [109]. Unfortunately, the

structure and patterns of sequence variation within the subspecies is not well characterized,

thereby perhaps underestimating the usefulness of intrasubspecific comparisons.

Studies published prior to the initiation of my dissertation analyzed a combined total

of six wild-derived strains and concluded that the levels of variation observed between wild-

derived and classical strains are significantly higher than among classical strains and that

the variation among wild-derived strains is distributed uniformly across the genome [101-

103]. However, these results were limited by the fact that a small fraction of the genome was

surveyed. Furthermore, a maximum of four wild-derived strains were analyzed in each study

yielding an inability to determine the level of intrasubspecific variation.

MOM and CASP are wild-derived inbred strains that were generated at the

Laboratory of Animal Genetics, Graduate School of Bioagricultural Sciences, Nagoya

University, Japan [74, 115]. The MOM inbred strain was derived by K. Kondo from wild *M.*

m. molossinus mice caught in Japan in 1972. The CASP inbred strain was derived from wild

24

M. m. castaneus mice caught in Los Baños, Philippines in 1994. The derivation of these strains is not thought to involve intercrossing with any other species, subspecies or strains.

c) Haplotype structure of inbred strains

Accurate understanding of the haplotype structure of inbred mouse strains may provide powerful approaches in the identification of molecular variants underlying quantitative trait loci [93, 97, 100-102, 116]. One approach to haplotype association mapping relies on associating phenotypic variation in inbred strains with their strain distribution patterns (SDP, the patterns of allelic similarities and differences among strains at a locus) [102, 117]. Because variants with different alleles may have the same SDP, the use of the latter simplifies the analysis of diallelic variation and is a staple in molecular phylogenetic studies. Therefore, the number, frequency and spatial distribution of SDPs are critical parameters to define structure of sequence variation. A recent study has shown that only 13 SDPs account for almost 99% of 1,465 variants identified in eight classical inbred strains over a 4.8 Mb region of mouse chromosome 1 [102]. Importantly, the authors report that, despite the small number of SDPs observed, the haplotypes of inbred strains are complex because variants with the same SDP are clustered together, but they do not generally occur in simple blocks. This study concluded that only a limited number of SDPs (on the order of the number of strains analyzed) are present in regions of similar sizes. However, it also noted that it is not known how the number of SDPs across the genome may vary depending on the number of strains analyzed.

Chapter 2 describes our findings regarding the level and patterns of genetic diversity, including insertions/deletions, among 20 *M. musculus* inbred strains, and two additional species of *Mus*, *M. spretus* and *M. spicilegus*. We specifically included two wild-derived strains of each of four *M. musculus* subspecies in order to provide a broader view of the genetic variation available in intersubspecific comparisons of inbred mouse strains, to gain

25

an estimate of the diversity present among wild populations of mice, and to characterize the

phylogenetic history of that variation. Our findings demonstrate the usefulness of wild-

derived inbred strains in characterizing and mapping genetic components of the DDK

syndrome and other complex phenotypes.

CHAPTER 2. Characterizing Genetic Diversity Among Wild-Derived Inbred Strains

The work described in this chapter was accomplished in collaboration with Dr. Elena de la Casa-Esperón, Timothy A. Bell, David A. Detwiler, Dr. Terry Magnuson, and Dr. Carmen Sapienza. The aim of these experiments was to define the level, spatial distribution, and phylogenetic history of the genetic variation in the *Mus* species using classical and wild-derived laboratory inbred strains. I contributed significantly to the sequencing efforts and data analysis used to identify and characterize genetic variants; and also to the writing of the manuscript. These results have been published in Genome Research [Ideraabdullah *et al.* 2004].

I. Frequency and distribution of sequence variants

We have determined the DNA sequence of 62 genomic segments located on 14 chromosomes. These segments include a single exon (both coding and UTRs) of 19 genes, seven fragments spanning 6.5 kb of the *Il9r* gene (≈66% of the entire gene) and 36 fragments spanning approximately 650 kb of the *Cctb6-Ap2b1* region of chromosome 11 (Figure 2.1). Genes were selected based of the following criteria; 1) candidate genes for schizophrenia and hypertension (*Agtr1a, Bdkrb2, Comt, Dao1, Diap1, Ncam1, Pparg,* and *Prodh*); 2) genes involved in DNA repair and cancer (*Mlh1,* and *Pms2*); 3) imprinted genes (*Igf2, Grb10,* and *Ube3a*); and 4) map location, to provide a wide overview of different genomic regions (*Clasp1, Rgs4, Dutp, Mecp2, Npr1,* and *Tgm1*). We amplified exons to increase the probability of successful amplification of specific products in all strains analyzed. Finally, multiple fragments in the *Il9r* and *Cctb6-Ap2b1* regions were sequenced in

order to characterize the spatial distribution of SDPs and the extent of haplotype sharing in wild-derived strains.
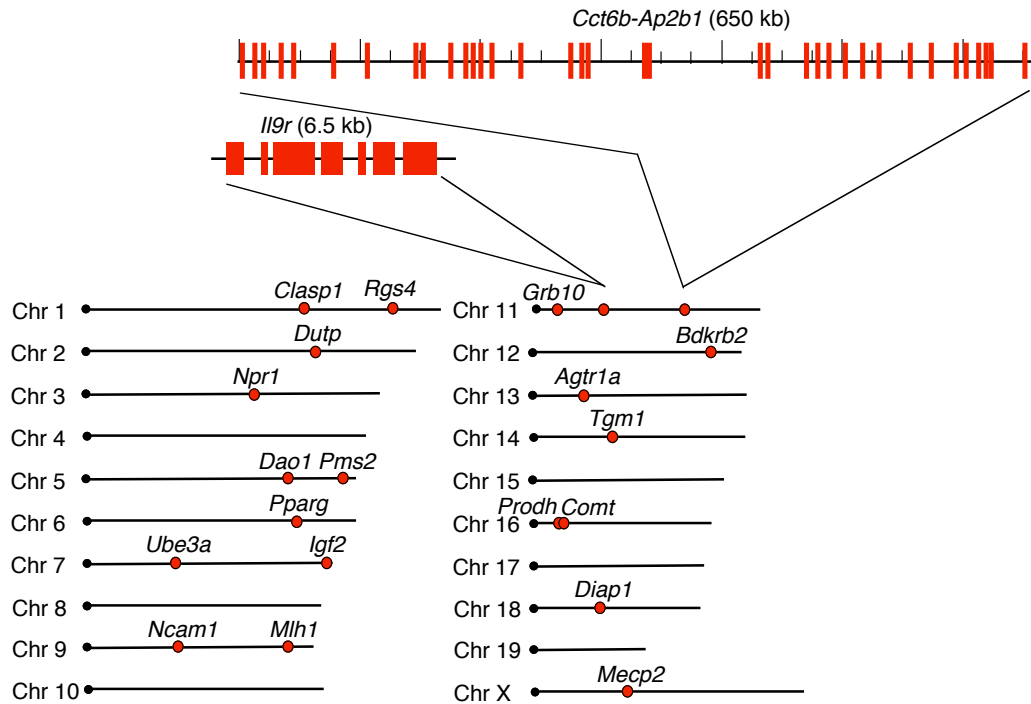
FIGURE 2.1. Genomic locations of fragments analyzed.

In total, 26,116 bp were sequenced from each strain. When all 22 strains are considered 1,007 sequence variants were identified, divided as follows: 89 microsatellite variants, 83 insertion/deletion variants and 835 substitution variants. Multiple alleles are found at most microsatellites, while insertion/deletion and substitution variants are largely diallelic (98.7 ± 0.4%). Therefore, we have omitted microsatellites from all subsequent analyses presented in this study. Table 2.1 shows the classification of variants by type and sequence context among inbred strains grouped according to their phylogenetic relationships. As expected, the estimated rates of variants per kb in this panel are substantially higher than previously reported in classical inbred strains [102]. Our estimates

TABLE 2.1. Classification of variants by context, type, and phylogenetic relationship.

Category	bp	Insertions/deletions			Substitutions		
		# of variants	Rate/kb	SE	# of variants	Rate/kb	SE
All *Mus* strains							
Intron/Intergenic	13154	57	4.33	0.57	487	37.02	1.65
UTR	6499	25	3.85	0.77	195	30.00	2.12
Coding	6463	1	0.15	0.15	153	23.67	1.89
M. musculus							
Intron/Intergenic	13154	35	2.66	0.45	314	23.87	1.33
UTR	6499	11	1.69	0.51	117	18.00	1.65
Coding	6463	0	0	0	105	16.25	1.57
M. m. domesticus							
Intron/Intergenic	13154	17	1.29	0.31	162	12.32	0.96
UTR	6499	6	0.92	0.38	42	6.46	0.99
Coding	6463	0	0	0	50	7.74	1.09
Classical							
Intron/Intergenic	13154	13	0.99	0.27	122	9.27	0.84
UTR	6499	2	0.31	0.22	28	4.31	0.81
Coding	6463	0	0	0	41	6.34	0.99

Analyses were performed among strains classified into the following four categories: 1) all inbred strains analyzed, 2) all M. musculus strains, 3) all wild-derived M. m. domesticus strains and 4) classical strains. The table shows the length of high quality sequence, the number of variants per kb and the SE of the rate. Variants were divided into insertions/deletions and substitutions. Discrimination between microsatellites and insertions/deletions was performed on the basis of the presence/absence of more than four adjacent tandem repeats of the inserted/deleted sequence, respectively. This threshold was determined empirically by comparing the frequency of mononucleotide runs of different sizes in the sequenced region and the frequency of insertions/deletions observed in runs of each size. For runs of less than five nucleotides, the probability of observing insertions/deletions is roughly equal to their frequency. For runs of five or more nucleotides, the probability of an insertion/deletion is ten to a hundred times greater than expected.

are also higher than the rates of variation observed in studies using fewer wild-derived

strains [97, 101, 102]. Overall, the highest density of variants is found in intron/intergenic

regions and the lowest in coding sequences (Table 2.1). Insertions/deletions display the

most pronounced difference in the rate of variation. The reversal in the density of

substitutions between coding sequences and UTRs in *M. m. domesticus* and classical

inbred strains (Table 2.1) most likely reflects limited sampling of regions that may have

different phylogenetic histories.

In pairwise comparisons, 35-45% of the total variation identified in our panel would

be observed in each one of the three possible types of inter-specific crosses between *M.*

musculus, *M. spretus* and *M. spicilegus* inbred strains (Figure 2.2). Classification of 918

variants according to the species in which the minor allele is present demonstrates that 63.4

± 1.6% (582/918 in Table 2.2) of the genetic diversity identified in our study is present

among subspecies of *M. musculus*. Because the panel includes only one *M. spretus*

(SPRET/EiJ) and one *M. spicilegus* strain (PANCEVO/EiJ), our study does not provide any

information about intra-specific variation in these two species. Interestingly, our analysis

identifies a subset of *M. musculus* diallelic variants in which the minor allele is present also

in either *M. spretus* or *M. spicilegus*, but not in both (Table 2.2). This class represents a

sizable fraction of the variants (8.2 ± 0.9%, 75/918 in Table 2.2) and raises questions

regarding the common assumption that variants identified in pairwise comparisons arise by

a mutation event in one of the two branches emerging from the last common ancestor. This

observation highlights the importance of comparisons using multiple samples of each taxon

and the usefulness of wild-derived strains to interpret molecular variation within a clear

evolutionary framework [109] (see below).

We have limited our analysis of intra-subspecific variation to the *M. m. domesticus*

subspecies, because the other three subspecies were represented in our panel by only two

strains each. We have estimated that a maximum of 58.4 ± 2.0% of the 582 variants

	SPRET	PANCEVO	CAST	CASA	CALB	JF1	MOLC	CZECHI	PWK	SKIVE	PERA	PERC	ZALENDE	TIRANO	LEWES	RBA	DDK	BALB/c	B6	DBA/2	A
	M. spretus	*M. spicilegus*	*M. m. castaneus*	*M. m. castaneus*	*M. m. castaneus x M. m. domesticus*	*M. m. molossinus*	*M. m. molossinus*	*M. m. musculus*	*M. m. musculus*	*M. m. musculus x M. m. domesticus*	*M. m. domesticus*	*M. m. domesticus*	*M. m. domesticus*	*M. m. domesticus*	*M. m. domesticus*	*M. m. domesticus*	Classical	Classical	Classical	Classical	Classical
PANCEVO	75																				
CAST	64	70																			
CASA	64	68	1,528																		
CALB	63	67	288	276																	
JF1	61	67	143	141	139																
MOLC	63	66	110	110	113	200															
CZECHI	61	67	121	118	123	183	200														
PWK	62	64	120	116	119	230	236	703													
SKIVE	61	63	103	99	102	160	168	236	304												
PERA	63	64	125	123	138	99	97	105	106	121											
PERC	70	69	122	118	133	100	107	105	110	103	167										
ZALENDE	65	67	122	120	136	102	105	111	113	144	196	237									
TIRANO	66	68	132	126	152	110	109	115	120	118	215	335	10,219								
LEWES	66	65	117	115	125	100	97	99	105	116	208	150	199	211							
RBA	63	63	125	125	144	104	96	104	109	103	173	258	171	198	191						
DDK	69	64	105	105	114	87	96	84	89	98	184	200	160	160	231	249					
BALB/c	64	65	114	113	123	95	96	97	102	120	225	166	198	186	249	344	355				
B6	63	64	122	123	138	105	101	108	115	108	172	185	179	183	191	500	270	461			
DBA/2	62	62	130	128	147	113	113	119	126	97	164	159	153	182	170	335	231	286	394		
A	62	63	144	147	152	118	111	132	134	106	307	182	165	194	148	233	174	195	243	349	
129X1	61	63	137	140	142	110	101	113	117	121	341	179	198	204	156	225	174	234	318	206	515

FIGURE 2.2. Estimated distance in base pairs between consecutive variants in pairwise comparisons.
The table provides the estimated distance between consecutive variants (including substitutions and insertions/deletions) in the 231 pairwise comparisons between the 22 strains analyzed.

`

identified in *M. musculus* (Table 2.1) would be present in three-way comparisons using a single strain from each of *M. m. castaneus*, *M. m. musculus* and *M. m. domesticus* subspecies. This fraction increases to 69.2 ± 1.9% with the inclusion of a *M. m. molossinus* strain. Table 2.1 indicates that over 47% of the total *M. musculus* variation is present in the *M. m. domesticus* subspecies (277/582 in Table 2.1). Therefore, over 50% of the total

TABLE 2.2. Classification of variants according to the distribution of the minor allele.

Minor allele present in	# of variants	% of total variants± SD	% of *M. musculus* variants± SD
SPRET	144	15.8 ± 1.2	n.a.
PANCEVO	121	13.2 ± 1.1	n.a.
SPRET & PANCEVO	71	7.8 ± 0.9	n.a.
M. musculus & SPRET	34	3.7 ± 0.6	5.8 ± 1.0
M. musculus & PANCEVO	41	4.5 ± 0.7	7.0 ± 1.1
M. musculus	507	55.5 ± 1.6	87.1 ± 1.4
Wild derived	301	32.9 ± 1.6	51.8 ± 2.1
Wild derived & Classical	185	20.2 ± 1.3	31.8 ± 1.9
Classical	21	2.9 ± 1.3	3.6 ± 0.8

The table shows 918 insertion/deletion and substitution variants identified in this study classified according to the distribution of the minor allele among three *Mus* species into six mutually exclusive classes. Variants found exclusively in *M. musculus* are further subdivided depending on whether the minor allele is found in wild-derived strains only, in classical strains only or in both types of strains. n.a., not applicable.

variation identified in *M. musculus* may be recovered in intra-subspecific crosses. On the other hand, only one third of the *M. musculus* variation is present in the six classical inbred strains analyzed here (Table 2.2). The variation observed in classical strains is slightly lower than the variation found in *M. m. domesticus* (Table 2.1), despite the fact that classical strains appear to have haplotypes derived from two different *M. musculus* subspecies in five out of the 62 fragments analyzed.

When all 20 *M. musculus* strains are considered, the density of variants across the

fragments analyzed follows the expectations under a random (Poisson) distribution,

suggesting that there is a uniform distribution of variants across the genome (Figure 2.3).

However, when comparisons are restricted to *M. m. domesticus* or classical inbred strains

there is an excess of fragments with no variants, as also observed by Yalcin and coworkers

(2004).
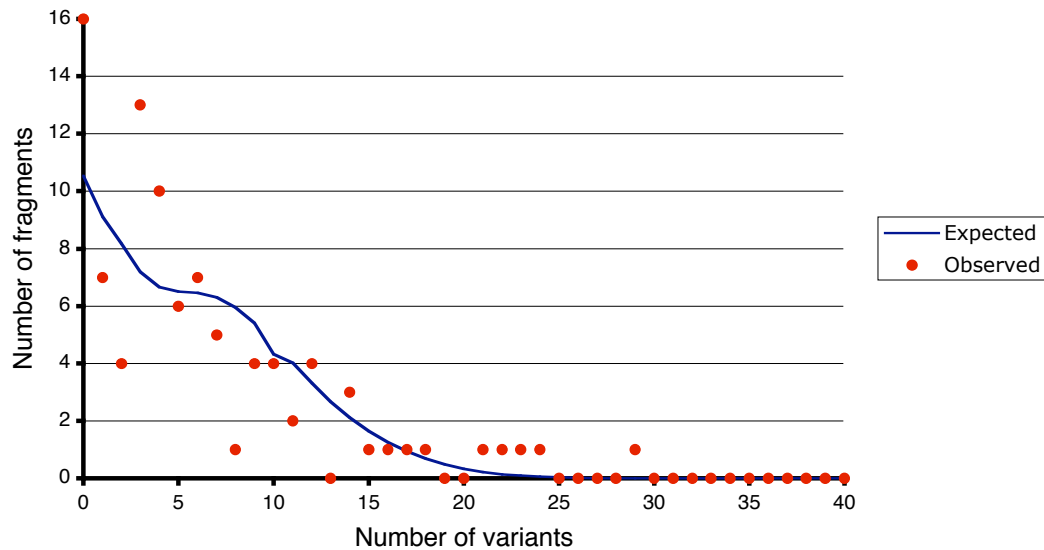
FIGURE 2.3. Expected and observed number of variants per fragment in *Mus musculus* inbred strains. The expected number of variants for fragments of different lengths is expected to follow a Poisson distribution and was calculated as described previously [102]. Briefly, expectations were calculated separately for each type of sequence based on the observed overall rates (Table 2.1). Overall rates were calculated by adding the observed substitution and insertion/deletion rates for each type of sequence context. The expected number of fragments with each number of variants was obtained by summing over all fragments.

II. Analysis of insertion/deletions

In total, we have identified 83 insertions/deletions representing approximately 9% of

the total variants (excluding microsatellites). The density of insertions/deletions varies in

different types of sequence. In introns/intergenic regions and UTRs, insertions/deletions represent 12% of the total variants while they are very rare (<1%) in coding regions (Table 2.1). The contribution of insertions/deletions to the diversity present in the 231 possible pairwise comparisons between the 22 strains analyzed here follows a normal distribution centered on the mean (data not shown). The size of the insertions/deletions ranges from 1 to 70 bp with an average of 5.4 bp (Figure 2.4). However, the distribution of insertions/deletions is strongly skewed towards smaller sizes. One bp insertions/deletions represent over 40% of the total variants and 80% of them are shorter than 6 bp (Figure 2.4). We were able to classify 81 of these 83 insertion/deletion variants as either insertions or deletions based on the predicted ancestral allele (identified using the SDP and the phylogenetic tree, see below). Deletions are significantly more frequent than insertions (52 *versus* 29, respectively; H_0: equal number of deletions and insertions, $\chi^2 = 6.53$, 1 df, $p<0.05$) and this trend is consistently observed in the three species analyzed here.
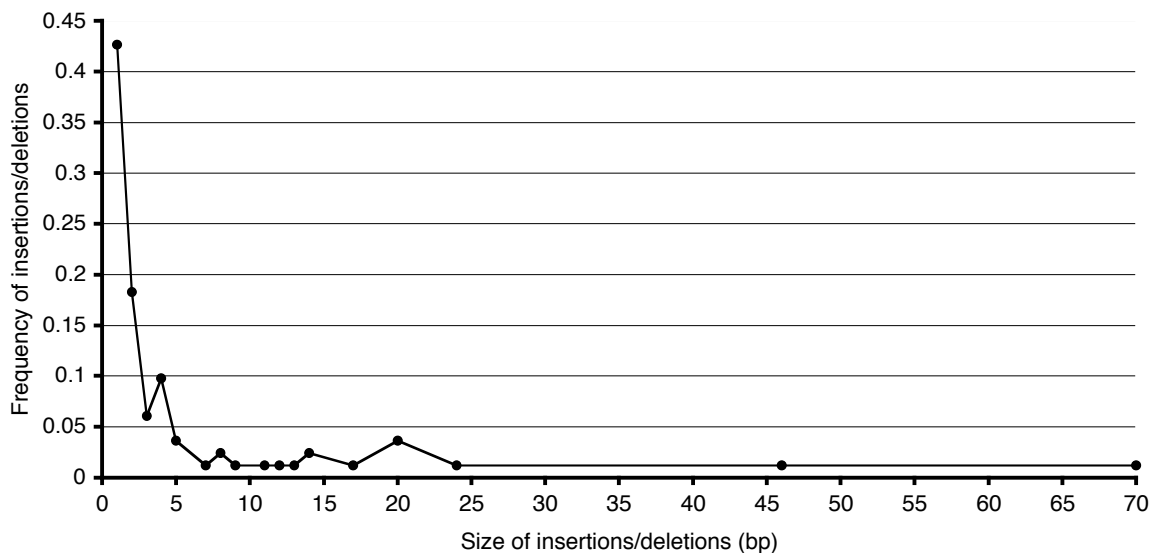
FIGURE 2.4. Frequency distribution of insertion/deletions sizes.

We have identified 164 SDPs in the 569 diallelic variants present among 20 *M. musculus* strains (variants found only in inter-specific comparisons and triallelic variants were excluded in this analysis). Both the number and frequency distribution of SDPs we observed differs from that observed previously. While similar numbers of SDPs (19) and strains (8) were noted by Yalcin and coworkers [102], we observe nearly an order of magnitude more SDPs than strains. In addition, 13 of the most common SDPs (those with a frequency greater than 1%) described by Yalcin and coworkers [102] accounted for 99% of the total variation. In our analysis, the 26 common variants (those with frequency >1%) represent only 57% of the total variation. Only 10 SDPs have frequencies >2% and more than half of the SDPs are defined on the basis of single variant (Fig. 2.4). The high number and low frequency of SDPs suggest that there is very limited haplotype sharing among the panel of *M. musculus* strains analyzed here.

If alleles at a locus are identical by descent, gene flow has not occurred between the different branches of the phylogenetic tree and there is no polymorphism at the branch points, the maximum number of SDPs that are consistent with any given phylogeny is $2n - 3$, were n is the number of strains. Therefore, for 20 strains the maximum number of SDPs is 37 and the 127 SDPs in excess of this number (Table 2.3) must be due to alleles that are identical by state (IBS) rather than identical by descent (IBD), the presence of ancient variants, and/or gene flow between the branches. The presence of gene flow is expected, given the fact that 10 out of the 20 strains included in our panel are intersubspecific hybrids and classical inbred strains (which also have a mixed phylogenetic history [110]). Alleles at variant positions at different regions of the genome in these strains may originate from different subspecies. In other words, the phylogenetic relationship between hybrid strains and classical strains varies at different loci. Therefore, exclusion of inter-subspecific hybrids

35

TABLE 2.3. Variation in number of SDPs observed at different genomic regions and in different sets of inbred strains.

Genomic region	Strains (#)	Observed # of SDPs	Excess # of SDPs	# of SDPs with variants segreg. across subspecies
Complete dataset	All (20)	164 (569)	127	n.a.
Complete dataset	Non hybrids (10)	60 (481)	43	26 (49)
Cct6b-Ap2b1	All (20)	83 (309)	46	n.a.
Cct6b-Ap2b1	Non hybrids (10)	44 (266)	27	18 (34)
Il9r	All (19)	41 (127)	6	n.a.
Il9r	Non hybrids (9)	17 (108)	2	4 (6)

Shown is the genomic region, the strains used in the analysis, the observed number of SDPs, the excess number of SDPs and the number of SDPs that segregate across *M. musculus* subspecies. Parenthesis show the number of variants used to define the SDPs. Excess number of SDPs is the difference between the observed number of SDPs and the maximum number of SDPs that may be observed in set of *n* strains if variants are identical by descent, there is no gene flow between branches and all variation occurred after the divergence of the subspecies. n.a., not applicable.

and classical strains should reduce the excess of SDPs, as should limiting the analysis to a single region of the genome. To fulfill these criteria, we compared the number SDPs present in both the total set of strains with the subset of 10 strains that are not known to be inter-subspecifc hybrids (CAST/EiJ; CASA/EiJ; CZECHI/EiJ; PWK/Ph; PERA/EiJ; PERC/EiJ; ZALENDE/EiJ; TIRANO/EiJ; LEWES/EiJ and RBA/DnJ). This analysis was performed independently in the total dataset, in the *Cct6b-Ap2b1* region (650 kb) and in the *Il9r* locus (6.5 kb). As expected, the number of excess SDPs in each group decreases significantly when only non-hybrid strains are analyzed (note that this decrease cannot be accounted for solely by the decrease in the number of variants, Table 2.3). Furthermore, the excess of SDPs decreases when the analysis is limited to smaller genomic regions. Therefore, the

mosaic nature of the genome in hybrid strains does contribute significantly to SDP diversity. However, in all cases there is still an excess of SDPs (Table 2.3) indicating that IBS and/or ancient polymorphisms are responsible. Inspection of SDPs in non-hybrid strains in each comparison demonstrates that alleles at 24-43% of the SDPs segregate across *M. musculus* subspecies. In approximately half of these SDPs, variant alleles segregate simultaneously in two different subspecies while in the other half each one of the two alleles found among *M. m. domesticus* inbred strains is found in either *M. m. musculus* or *M. m. castaneus* but not both. We conclude that the presence of a large fraction of variants that appear to segregate across subspecies contribute to the large number of SDPs found in our panel. The remaining excess is probably due to gene flow within subspecies rather than IBS (see below).

IV. Phylogenetic history of the genetic variation found in *Mus musculus*

Two of our previous analyses suggest that a considerable fraction of the total *M. musculus* variants segregate across species (Table 2.2) and across subspecies (Table 2.3). To formally address when the mutation events took place and which is the ancestral allele at each variant position, we determined the most parsimonious way to explain the SDPs that is also consistent with the true phylogeny (see Methods). The 569 diallelic variants, including insertions/deletions and substitutions, found among classical and wild-derived strains of *M. musculus* can be classified into three categories on the basis of whether the mutation event occurred before the divergence of the three specific lineages (Figure 2.5a), before the divergence the *M. musculus* subspecies (Figure 2.5b) or after the divergence of the *M. musculus* subspecies (Figure 2.5c). Interestingly, 37.8 ± 2.0% of the total variants appear to predate the divergence of the *M. musculus* subspecies (Figure 2.5a & b). These "ancient" polymorphisms are distributed uniformly across the fragments sequenced. Importantly, on average, they represent 44.7 ± 6.7% of the sequence variants observed in pairwise

37

comparisons between *M. musculus* strains (Figure 2.6). In other words, the contribution of

ancient polymorphisms to the sequence diversity is higher in pairwise comparisons than in

the total data set, due to the higher frequency of the minor allele in this type of variant.

	a	b	c
Mus spretus			
Mus spicilegus			
M. m. domesticus			
M. m. musculus			
M. m. castaneus			
# of variant events	75	140	354
Total		215	354
# of transitions		145	228
# of transversions		53	99
# of insertions/deletions		17	27

$\chi^2 = 0.74$

FIGURE 2.5. Phylogenetic history of *Mus musculus* variants. The figure shows a simplified phylogenetic tree of the *Mus* lineage [109]. Insertion/deletion and substitution variants were assigned to three different branches of the tree (see Chapter 6). Variants were further classified as transitions, transversions and insertions/deletions. The chi-square value was calculated based on a test for independence between the origin of the variant and the type of substitution.

Because our ability to assign some *M. musculus* variants to the branch connecting

the specific and subspecific divergence nodes (Fig. 2.5, b) depends on the correct

identification of the ancestral allele, we tested whether our predictions are supported by the

allele found at homologous positions in the rat. The predicted ancestral allele is supported

by the rat sequence at 82.4% of variants at which rat has one of the two alleles found in *M.

musculus*. On the other hand, we observe significant statistical evidence of an increase in

the ratio of transitions to transversions at the remaining 17.6% of variants (H_0= the ratio of

transitions and transversions is independent of the whether the predicted ancestral allele is

consistent with the rat allele, 1 d.f., χ^2= 4.74; p<0.05). This observation indicates that two

substitution events (*i.e.,* IBS) , one in the *Mus* lineage and another in the rat lineage, are
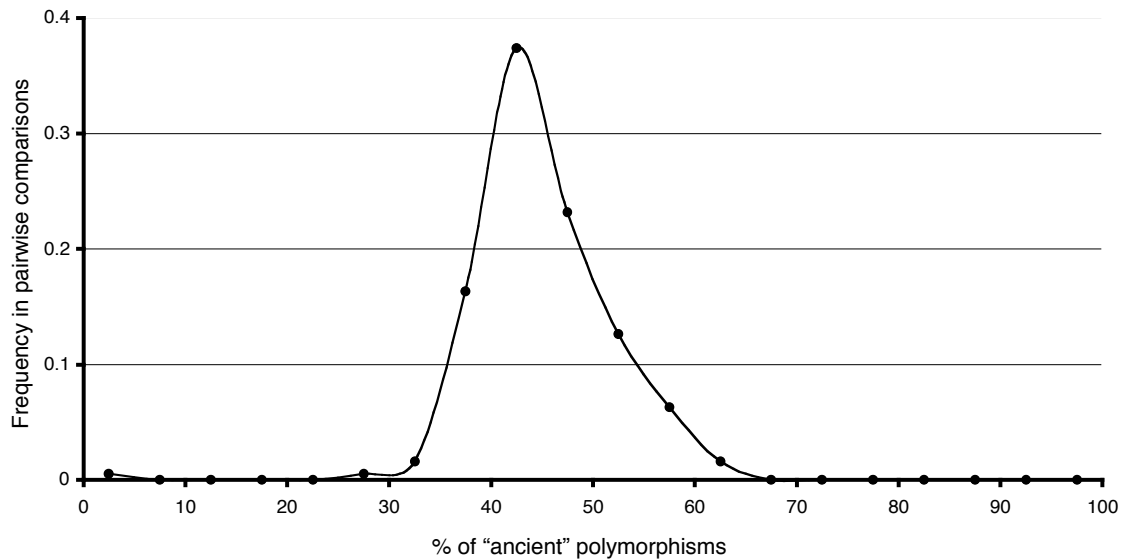
FIGURE 2.6. Frequency distribution of ancient polymorphisms in pairwise comparisons between *M. musculus* inbred strains. The percent of ancient polymorphisms was calculated in each of 190 pairwise comparisons between the 20 *M. musculus* inbred strains analyzed in our panel. Frequencies in pairwise comparisons were assigned to one of 20 equal percentile classes with respect to the percent of ancient polymorphism.

responsible for the inconsistencies. We conclude that errors in the determination of the

ancestral allele are small and should not affect significantly our classification of variants

segregating among *M. musculus* strains.

V. Discussion

The level of genetic diversity reported here is higher than in previous reports [97,

100-102]. This holds true for the levels of genetic diversity observed in pairwise

comparisons (Figure 2.2), the estimated rate of variants (Table 2.1) and the number and

diversity in SDPs (Figure 2.7). Regional differences in sequence variation, similar to those

reported among genomic regions in comparisons between human and chimpanzee [118],

may be partly responsible for the differences. We suspect that sampling of a larger

collection of strains with greater ancestral diversity account for most of the discrepancy

between our results and those of previous reports.
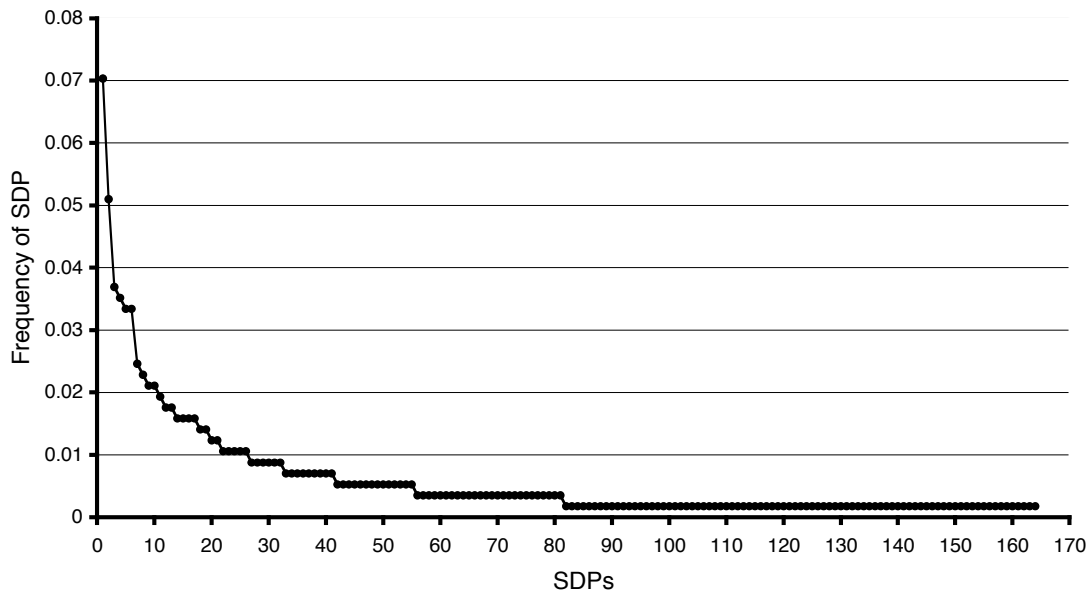
FIGURE 2.7. Frequency distribution of SDPs. The frequency at which each one of 164
SDPs is observed among 569 *M. musculus* dialleic variants is shown in descending order.

To appreciate the full extent of the genetic diversity in *M. musculus*, it is instructive to

compare our results with the sequence diversity observed among and between closely

related mammalian species. Our estimate of the average frequency of SNPs in

intersubspecific crosses in mouse (one SNP per 0.11 kb) is ten times higher than in humans

and three times higher than in chimpanzees (one SNP per 1.08 kb and one SNP per 0.38

kb, respectively) [119-123]. The estimate in chimpanzee has been corrected to account for

the fact that the region analyzed is in the X chromosome and has low mutation rates [124].

In addition, the percentage of variant positions in the mouse genome, excluding

microsatellites, (3.1%) is one order of magnitude higher than in humans (<1%) and between

two and thirty times higher than in chimpanzees, bonobos, gorillas and orangutans (1.7%,

0.1%, 0.8% and 1.5%, respectively) [124]. Although the existence of distinct subspecies

contributes to the high level of diversity observed in the mouse, the analyses of both

chimpanzee and orangutan also included individuals from distinct subspecies. The shorter

generation time in mouse and the fact that a substantial fraction of the variation predates the

divergence of the subspecies contribute significantly to the diversity. We conclude that the

mouse is the mammalian species with the highest levels of genetic diversity yet described.

In fact, the level of sequence diversity observed in *M. musculus* is more similar to that found

between man and chimpanzee than between individual humans [118, 125, 126].

Although our estimate for the sequence variants in the *M. musculus* may appear high

(Table 2.1), it represents, in all likelihood, an underestimate of the true value due to the limited

sample size, the presence of sampling biases and the type of variation detected. Mouse is a

polytypic species with five to six recognized subspecies. Two models have been proposed to

explain the origin and radiation of the commensal mice. The centrifugal model proposes that

mice radiated outward from the central populations found on the Indian subcontinent [127-129].

The linear model proposes that mice originated in West Eurasia and spread easterly to give

rise to the progenitors of the different subspecies [129]. Regardless of the model,

representatives of the central populations in the centrifugal model (or representatives of *M. m.*

castaneus in the linear model) harbor a significant fraction of the genetic variation of the whole

species [129]. The absence of strains from this group and from the *M. m. gentilulus* subspecies

in our study would lead to underestimation of the true level of genetic diversity in the mouse.

Furthermore, our analyses indicate that strain sampling in *M. m. castaneus* and *M. m.*

musculus subspecies may itself be biased because relatively little diversity has been captured

among the strains of these two subspecies. Given the range of the distributions shown in other

inter-subspecific comparisons (Figure 2.2), this circumstance is most likely due to the limited

geographical origins of these strains [130] rather than an inherent lack of diversity within some

subspecies. These data suggest that the fraction of variant positions that are present in mouse may be much higher than our estimate.

Here we provide the first estimate of the level of variation among inbred strains of *M. musculus* subspecies (previous studies [97, 100-102] included only one wild derived representative of four *M. musculus* subspecies and classical strains. Our analyses of six wild-derived *M. m. domesticus* indicate that there is considerable variation within a subspecies (Table 2.1). In fact, in some fragments the level of variation between *M. m. domesticus* inbred strains is similar to the average diversity found in inter-subspecific pairwise comparisons (Figure 2.2). Although previous studies in classical inbred strains have equated the presence of segments with high frequency of polymorphisms with different subspecific origin [97, 100-102], the high level of intrasubspecific diversity shown here demonstrates this is not a requirement to observe high levels of sequence variation (Figure 2.8). It also suggests that some wild-derived inbred strains descend from branches diverging early within subspecific lineages.

The analysis of SDPs in our panel of inbred strains complements the study of Yalcin and coworkers [102], but also shows some striking contrasts, including the higher number and prevalence of rare SDPs in our panel (Figure 2.7). Although these studies differ in the number and type of strains and the location of the sequence analyzed, some useful comparisons can be made. When genome-wide SDP analysis in our panel is restricted to the six classical strains, the number and frequency distribution of SDPs are almost identical (data not shown) to those reported previously for eight strains in a 4.8 Mb region of chromosome 1 [102], suggesting that only a limited number of SDPs may be present in small sets of classical inbred strains. On the other hand, significant increases in SDP number may be achieved with the inclusion of wild-derived strains represented. Whether there is a predictable number of genome-wide SDPs in a panel remains unknown. However,
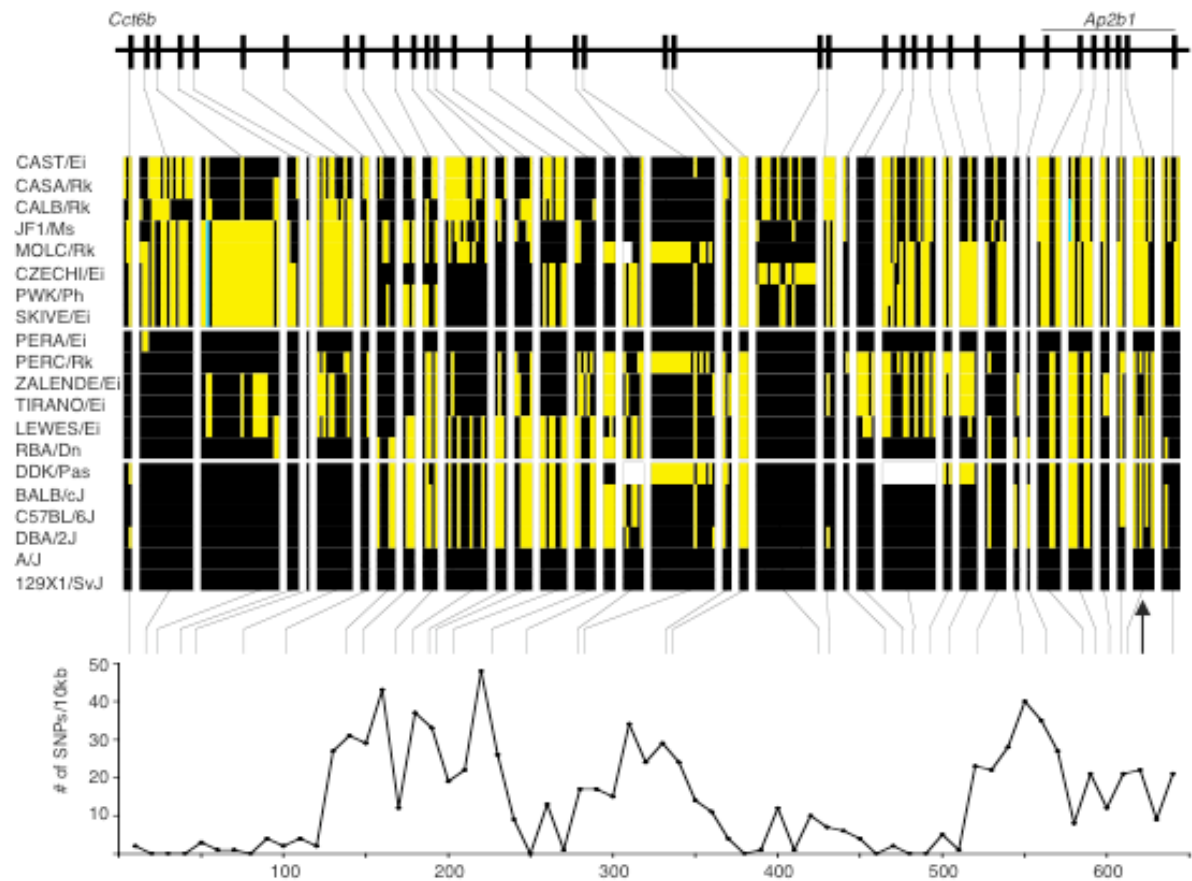
FIGURE 2.8. Haplotype diversity in *Mus musculus*. (a) Vertical bars indicate the locations of the 36 sequenced fragments in the 650 kb *Cctb6-Ap2b1* interval of chromosome 11 (the fine horizontal line spans the fragments sequenced within the latter gene). (b) Haplotypes are based on the alleles present at the positions of 263 substitution and insertion/deletion variants that remain after eliminating variation in microsatellites and variants found in a single strain. Each variant is represented as a vertical bar of equal width. At diallelic sites, the two alternative alleles are shown as yellow and black bars. The presence of a third allele at two positions is shown in green. Contiguous variants were used to define the haplotype in each fragment. Consecutive fragments are separated by vertical open spaces. Strains have been assigned to one of three groups; "classical" (bottom), wild-derived *M. m. domesticus* (middle), and wild-derived non-*M. m. domesticus* (top). Groups are separated by horizontal open spaces. The arrow designates the fragment showing putative gene conversion events. The two white boxes in the DDK/Pas strain indicate fragments that fail to amplify using the original primers (c) Analysis of SNP density. All SNPs in the Mouse RefSNP section of the Celera Discovery System were used with the exception of positions at which the putative SNPs was supported only by the presence of an intra-strain polymorphism. The number of SNPs per 10 kb interval was calculated and is indicated in the vertical axis. The horizontal axis shows distances in kb.

it is evident that very large number of SDPs may be found in panels of strains that include either natural or artificial hybrids. Furthermore, increase in SDP diversity not only depends on the number and type of strains, but also on the pervasive presence of ancient polymorphisms. This SDP diversity should provide higher mapping resolution but decrease the statistical power [102]. Whether this trade off is suitable will depend on the type of genetic experiment to be conducted.

Our results confirm that the taxonomical classification of some wild-derived strains may need revision as sequence data accumulate. For example, although CALB has been assigned to the *M. m. domesticus* subspecies it has haplotypes related to *M. m. castaneus* in at least five regions located in three different chromosomes, including the *Il9r* gene and the *Cct6b-Ap2b1* region. These results support the idea that CALB is a hybrid of *M. m. domesticus* and *M. m. castaneus.* This raises the possibility that some inbred strains, derived from animals trapped in the periphery of the range of a subspecies or from recently colonized areas, may have a mosaic genome. This may be especially relevant in the case of the *M. m. domesticus* subspecies*,* as some of the more easily available and most commonly used wild-derived strains from this taxon are derived from mice trapped in the Americas (*i.e.*, WSB, PERA, PERC, LEWES, etc).

We also wish to note that "ancient" polymorphisms represent a substantial fraction of the total variants (Figure 2.5). As stated in the previous section, three mechanisms may explain the presence of such polymorphisms in *M. musculus*. First, the variants may reflect gene flow among species and subspecies. Introgression of genetic variants between *M. spretus* and *M. musculus* has been reported [131, 132]. However, the high frequency of such events in our data (13.2 $\pm$ 1.4%, Figure 2.5) and the almost uniform distribution of such events across the regions analyzed (data not shown) does not support recent inter-specific gene flow. In the same vein, introgression of genetic variants between subspecies is

common in hybrid zones and in classical strains [109], but our approach specifically excludes the use of strains derived from hybrid populations in the identification of ancient polymorphisms (Methods). Therefore, gene flow is unlikely to contribute significantly in explaining the level of the ancient polymorphisms reported here. On the other hand, these variants may be examples of polymorphisms that have been maintained through long evolutionary periods or, alternatively, IBS (*i.e.,* the reoccurrence of the same mutation event in different lineages). The ratio of transitions to transversions may be used to discriminate between substitution variants IBD and IBS. In the case of IBS, the ratio of transitions to transversions should increase by twice the ratio of transitions to transversions in IBD. We found no statistical evidence for an excess of transitions among variants predating the divergence of the subspecies when compared to variants arising after it (Figure 2.5).

Identification of ancient variants depends on the number and phylogenetic relationship of the samples analyzed. In this report, most ancient variants would have gone unrecognized without the inclusion of representatives of both *M. spretus* and *M. spicilegus* and of multiple inbred strains from each of three *M. musculus* subspecies. Our analysis relies on the use of the phylogenetic tree proposed by the centrifugal model [127, 128]. Some of our conclusions will require reevaluation if the linear model ultimately represents the true phylogeny of *M. musculus* subspecies. However, this reinforces the need for the characterization of genetic diversity in a wide and representative panel of wild-derived inbred strains. The persistence of ancient polymorphisms may help to explain how inter-fertile populations of the same species have maintained a greater degree of sequence diversity than that found between man and chimpanzee. Although the data presented here are based on inbred strains, our observations suggest that the effective population size in the *Mus* lineage has been relatively large and constant over a long evolutionary period.

These variants contribute significantly the genetic diversity present in *M. musculus*, but they may also, if unrecognized, affect the conclusions of evolutionary and haplotype studies.

Finally, our analysis of insertion/deletion variants provides an example of the value of wild-derived strains to interpret the molecular variation within a clear evolutionary framework [109]. The size of the mouse genome is significantly smaller than in humans [133]. Based on genome-wide comparisons between mouse and human genome it has been proposed that the smaller size in the mouse is not simply the result of an increase in genome size in the human lineage (due to duplications and the addition of repetitive elements) but to the loss of ancestral sequences in the mouse lineage [133]. Our analysis of just over 80 insertion/deletion variants confirms this conclusion and suggests that it is an ongoing process in three different *Mus* species. Further studies are required to determine the relative contribution of small deletions, such as those reported here, to the overall decrease in size in the mouse genome.

CHAPTER 3. High Resolution Mapping of the Paternal Gene of the DDK Syndrome

The work described in this chapter was accomplished in collaboration with Timothy

A. Bell, Dr. Elena de la Casa-Esperón, Heather E. Doherty, Kuikwon Kim, Dr. Yunfei Wang,

Dr. Leslie A. Lange, Dr. Kirk Wilhemsen, Dr. Ethan M. Lange and Dr. Carmen Sapienza.

The aim of these experiments was to utilize naturally occurring genetic variation among

inbred strains in recombination and association mapping to significantly reduce the size of

the candidate interval for the paternal gene at *Om*. I contributed significantly to this study in

the collection and analysis of data used to determine the reproductive performance of inbred

males, which was then used in the *in sillico* analysis to map the paternal gene. I also

assisted in the writing of the manuscript describing these results, which was published in

Genetics [Bell *et al.* 2006].

I. Mapping strategies

Mapping the paternal component of the syndrome requires that the females used in the

experimental crosses produce oocytes with the maternal DDK factor. If this requirement is

fulfilled, mapping the paternal gene is straightforward because the fate of the resulting

embryos depends exclusively on whether they inherit a compatible or incompatible allele at

the paternal gene [66]. Therefore, there is an inverse proportional relationship between the

number of incompatible alleles (0, 1 or 2) in a male and its reproductive performance (see

Chapter 1, Table 1.1). In addition, the presence of transmission ratio distortion against

incompatible alleles (see Chapter 1, Table 1.1) can be used to determine whether males of

mixed genetic background are homozygous or heterozygous for compatible and

incompatible alleles at the paternal gene. In this study we use litter size as an indirect measure of the level of embryonic lethality [69, 75]. In crosses involving (B6 x DDK)F_1 hybrids females the litter size is normally distributed around the mean and depends on the genotype of the sire at the paternal gene [69, 75]. In contrast, the litter size is not normally distributed in crosses involving DDK females and most matings do not produce live pups [75]. In these crosses, characterization of the DDK syndrome phenotype requires using both litter size and delivery ratio data and is subject to greater uncertainty [75].

II. Recombination mapping

We have characterized the reproductive performance of crosses between identical (B6 x DDK)F_1 females and 40 males with mixed B6 - DDK background (Table 3.1). Crosses yielded an average of 29 litters (range: 10-47). The males can be divided into two groups based on their recombination status in the 1.5 Mb interval defined previously by Baldacci and coworkers [68, 70]. Twenty-two males carry non-recombinant chromosomes in this interval (Table 3.1a) while the remaining 18 males carry different combinations of five types of recombinant chromosomes defining six smaller intervals (Figure 3.1 and Table 3.1b).

First we confirmed that the 1.5 Mb interval analyzed contains the paternal gene using the reproductive performance of the 22 males with non-recombinant chromosomes (Figure 3.1). Males homozygous for DDK alleles have the best reproductive performance (average litter size = 9.48 ± 0.59), males homozygous for B6 alleles have the worst reproductive performance (average litter size = 4.51 ± 1.96) and heterozygous males have an intermediate behavior (average litter size = 7.06 ± 0.51). These averages are as expected from mean and distribution of the litter sizes of the control crosses [69, 75].

III. Association mapping using recombinant males

We then determined which of the six smaller intervals contains the paternal gene by testing for association between the genotypes present in these intervals and the

48

TABLE 3.1. Genotypes at selected markers in the vicinity of *Om* and reproductive performance of males used for defining the candidate interval.

a. Non-recombinants

Male	*D11Mit33* (81387805)	*Scya7* (81660424)	*D11Spn1* (81773654)	*D11Mit66* (82307550)	*D11Mit283* (82370768)	*D11Spn173* (82560436)	*D11Spn178* (82585484)	*D11Spn128* (82970687)	*D11Spn129* (83025035)	*Scya5* (83131570)	*D11Mit120* (83192071)	*D11Mit35* (83252491)	Average Litter Size	SD	# of Litters
531C	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	4.51	1.96	43
337D	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	5.94	2.04	18
502C	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	6.53	1.91	45
363L	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	6.86	2.25	29
891C	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	6.87	1.60	47
849C	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	7.05	1.84	44
719C	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	7.13	2.06	16
885C	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	7.27	1.99	33
516D	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	7.27	2.61	44
430L	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	7.45	2.14	20
854D	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	7.59	1.53	22
848C	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	BK	7.77	1.42	26
569D	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	8.70	3.27	37
954D	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	8.82	1.29	17
434H	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	9.00	2.08	30
249L	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	9.21	2.53	29
501C	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	9.27	1.70	45
431E	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	9.42	2.15	45
40E	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	9.67	1.81	33
534C	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	9.96	1.06	27
957D	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	10.27	1.39	22
430E	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	10.40	1.57	47

TABLE 3.1
b) Recombinants

Male	D11Mit33 (81387805)	Scya7 (81660424)	D11Spn1 (81773654)	D11Mit66 (82307550)	D11Mit283 (82370768)	D11Spn173 (82560436)	D11Spn178 (82585484)	D11Spn128 (82970687)	D11Spn129 (83025035)	Scya5 (83131570)	D11Mit120 (83192071)	D11Mit35 (83252491)	Average Litter Size	SD	# of Litters
969X	KK	KK	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	3.80	1.70	15
153T	KK	KK	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	3.93	1.86	14
152T	KK	KK	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB	4.33	1.54	15
115AA	BK	BK	BK	BK	BK	BK	BB	BB	BB	BB	BB	BB	4.71	2.43	14
537S	BK	BK	BB	BB	BK	BK	BK	BK	BK	BK	BK	BK	5.44	2.15	27
535S	BK	BK	BB	BB	BK	BK	BK	BK	BK	BK	BK	BK	6.44	2.04	27
534S	BK	BK	BB	BB	BK	BK	BK	BK	BK	BK	BK	BK	6.79	2.25	24
533S	BK	BK	BB	BB	BK	BK	BK	BK	BK	BK	BK	BK	6.90	2.09	31
536S	BK	BK	BB	BB	BK	BK	BK	BK	BK	BK	BK	BK	7.00	1.98	34
537-1S	BK	BK	BB	BB	BK	BK	BK	BK	BK	BK	BK	BK	7.25	2.29	32
735L	KK	KK	KK	KK	KK	KK	BK	BK	BK	BK	BK	BK	7.67	1.93	33
840X	KK	KK	KK	KK	KK	KK	KK	KK	BB	BB	BB	BB	8.57	1.63	21
323X	BB	BB	BB	BB	KK	KK	KK	KK	KK	KK	KK	KK	8.59	1.99	27
561T	BK	BK	BK	BK	KK	KK	KK	KK	BK	BK	BK	BK	8.84	2.77	31
844X	KK	KK	KK	KK	KK	KK	KK	KK	BB	BB	BB	BB	8.97	2.68	32
559T	BK	BK	BK	BK	KK	KK	KK	KK	BK	BK	BK	BK	8.91	2.52	32
560T	BK	BK	BK	BK	KK	KK	KK	KK	BK	BK	BK	BK	9.40	2.07	10
804L	KK	KK	KK	KK	KK	KK	KK	KK	KK	KK	BK	BK	10.30	1.78	20

In each class, males are sorted according to increasing average litter sizes.
Alleles are listed as K, DDK; or B, B6.

reproductive performance of both the complete set of 40 males and the 18 recombinant

males. Only two intervals, *D11Spn178-D11Spn128* and *D11Mit283-D11Spn173*, show

significant association in both the complete set of males and the recombinant males only

(Figure 3.1). The most significant values were obtained for the *D11Spn178-D11Spn128*

interval. Again, males homozygous for DDK alleles in this interval have the best
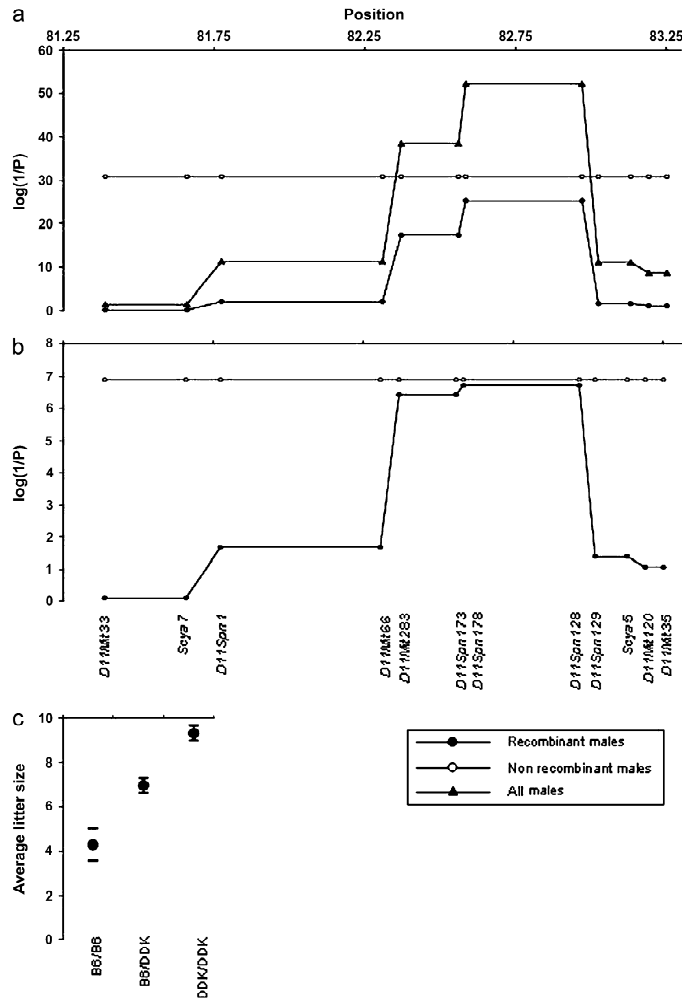
FIGURE 3.1. Mapping by recombinant progeny testing in B6-DDK background. The reproductive performances of 40 males were used to map the paternal gene in the candidate interval defined by Baldacci and coworkers [68, 70]. The original interval was divided into six smaller intervals based on the presence of five recombinant chromosomes. The markers shown in the bottom of section b convey the complete genotypic information (Table 3.1). In sections a and b the horizontal axes show distance in Mb from the centromere of chromosome 11. The vertical axes show the statistical significance of the association findings. Triangles show the results using the entire set of 40 males. Open circles show the results for the 22 males carrying non-recombinant chromosomes in the *D11Mit33 – D11Mit35* interval. Filled circles show the results in the 18 males carrying recombinant chromosomes in the *D11Mit33 – D11Mit35* interval (the inset in the bottom right shows this information visually). a) The top plot shows the strength of the associations between genotype and litter size using the F-distribution to assess the statistical significance of the mixed models. b) The bottom plot shows the strength of the association between genotype and litter size using permutation tests to assess the statistical significance of the mixed models. Empirical results for the 22 non-recombinant males and for the regions containing markers *D11Spn173* and *11Spn178* among the 18 recombinant males were determined exactly. The statistical significance when analyzing all 40 males for the regions containing marker *D11Mit33* (*p*=0.045) was estimated using a permutation test while the statistical significance for *D11Spn173* (*p*=1.3 x 10 -15) and *D11Spn178* (*p*=3.4 x 10-16) were calculated exactly (data not shown). It was not feasible to calculate the empirical statistical significance for the other regions (*p*<1.0 x 10 -5) using all 40 males. The statistical significance of the findings in these other regions were beyond the resolution of 100,000 random replications of the data and there were too many permutations of the data that would yield a more extreme value than the observed F-statistics to make it possible to determine the empirical significance exactly. c) Circles represent the average litter size in males with the three possible genotypes in the *D11Spn178-D11Spn128*. The horizontal bars show the boundaries of the 99% confidence intervals.

reproductive performance (average litter size = 9.07 ± 0.60), males homozygous for B6

alleles have the worst reproductive performance (average litter size = 4.19 ± 0.41) and

heterozygous males have an intermediate behavior (average litter size = 6.81 ± 0.71).

These results indicate that the paternal gene lies between the two recombinations defining

this interval (Figure 3.1).

This conclusion was confirmed by analyzing the inheritance of paternal alleles in the

surviving offspring of the two males, 115AA and 735L (Figure 3.2 and Table 3.1b), that have

different genotypes in the *D11Spn178-D11Spn128* and *D11Mit283-D11Spn173* intervals.

Each of these males is heterozygous for one interval and homozygous for the other.

Because the DDK syndrome embryonic lethality requires inheritance of a paternal B6 allele,

FIGURE 3.2. Transmission ratio of paternal haplotypes in the *Om* region from sires with critical recombinant chromosomes to their progeny. The figure shows the haplotypes in the vicinity of *Om* of the two males, 115AA and 735L, that carry recombinant chromosomes between *D11Spn173* and *D11Spn178*, and the number of offspring inheriting each paternal haplotype. The figure also provides the level of significance for each interval using the chi-square test statistics under the null hypotheses of equal transmission of alleles in the progeny of homozygous males for that interval and 66% transmission of the DDK allele in the progeny of heterozygous males for that interval [56]. n.s., not significant.

52

there should be an excess of offspring inheriting paternal DDK alleles in the progeny of the male that is heterozygous for the interval containing the paternal gene; and equal inheritance of paternal alleles in the male that is homozygous for the interval containing the paternal gene. As shown in Figure 3.2 there is a highly significant transmission ratio distortion against the inheritance of the recombinant chromosome among the offspring the 735L male while there is equal transmission of both chromosomes in the progeny of the 115AA male. These results are in agreement with the paternal gene lying in the *D11Spn178-D11Spn128* interval while the proximal interval may be rejected (Figure 3.2).

In conclusion, we have defined a 465 kb interval that contains the paternal gene responsible for the DDK syndrome. This interval is flanked by the excluded markers *D11Spn173* and *D11Spn129* (Figure 3.1).

IV. Characterization of the paternal gene compatibility of various inbred strains

We have characterized the reproductive performance of (B6 x DDK)F_1 females mated to males from 15 inbred strains (11 wild-derived strains: PANCEVO, CAST, JF1, MOLC, SKIVE, PERA, PERC, ZALENDE, TIRANO, LEWES, and RBA and four classical strains: A, 129X1, DBA/2 and BALB/c). Crosses yielded an average of 47 litters (Range: 16-103). We have corrected the litter size in inter-specific and inter-subspecific crosses to account for any reductions in litter size (as compared to intra-subspecific crosses) that are unrelated to the DDK syndrome (see Chapter 6). Figure 3.3 shows the mean litter size and 99% confidence intervals for these crosses. The figure also includes the reproductive performance of B6 and DDK males obtained in previous studies [55, 69, 75]. Inspection of Figure 3.3 reveals that there is considerable variation in the embryonic lethal phenotype depending on the inbred male used in the cross.

Average litter size
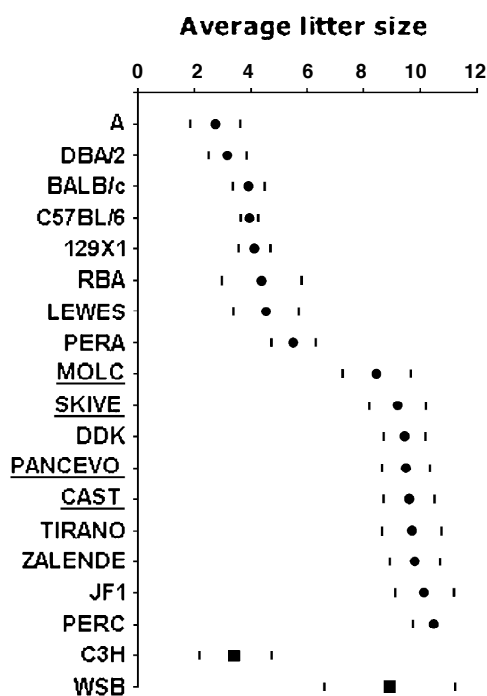
FIGURE 3.3. Reproductive performance of males from different inbred strains.
The vertical axis represents the average litter size observed in crosses between (B6 x DDK)F_1 females and males from the inbred strain listed in the horizontal axis. Circles denote the strains used for in silico mapping and squares are the strains used to confirm the presence of complete linkage disequilibrium between the phenotype and selected SNPs. The horizontal bars show the boundaries of the 99% confidence intervals adjusted for the correlation between litters from the same sire. Correction factors were applied to determine the reproductive performance of the underlined strains (see Chapter 6).

V. Association mapping using inbred strains

To further refine the location of the paternal gene we tested whether the reproductive

performance (using the average litter size as a quantitative trait, see Chapter 6) of the 17

inbred strains analyzed in the previous section is associated with the alleles present at any

of 167 variants distributed across 26 regions spanning the entire *D11Spn173-D11Spn129*

candidate interval defined by recombinant progeny testing. The results of this analysis are

shown in Figure 3.4. Several variants are significantly associated with reproductive

performance. These 16 variants are distributed across four fragments spanning 128 kb

(Figure 3.5a). Two diallelic SNPs at positions 82,843,176 and 82,843,476 (NCBI Build 33)

show the strongest association with the phenotype. Compatible strains carry G and T

alleles, respectively, at these two SNPs while incompatible strains carry the A and G alleles

(SNPs 10 and 11 in Table 3.2). The level of significance is exceptionally high and these

variants are optimally associated with the reproductive performance phenotype (*i.e.*, given

54

the available phenotype data, it is not possible to construct any combination of genotypes

for a SNP across the 17 strains that could better explain the phenotype than what we have

observed for these two variants). To test whether the alleles at these two SNPs predict the

DDK syndrome phenotype in other mouse strains, we genotyped four additional strains with

known phenotype (C3H, KK, CBA and PWK) and determined the phenotype of a strain of

known genotype (WSB). In each case the alleles present at these two SNPs predicted

accurately the compatible/incompatible phenotype of the strain, confirming strong

association with the paternal gene (Table 3.2 and Figure 3.3).
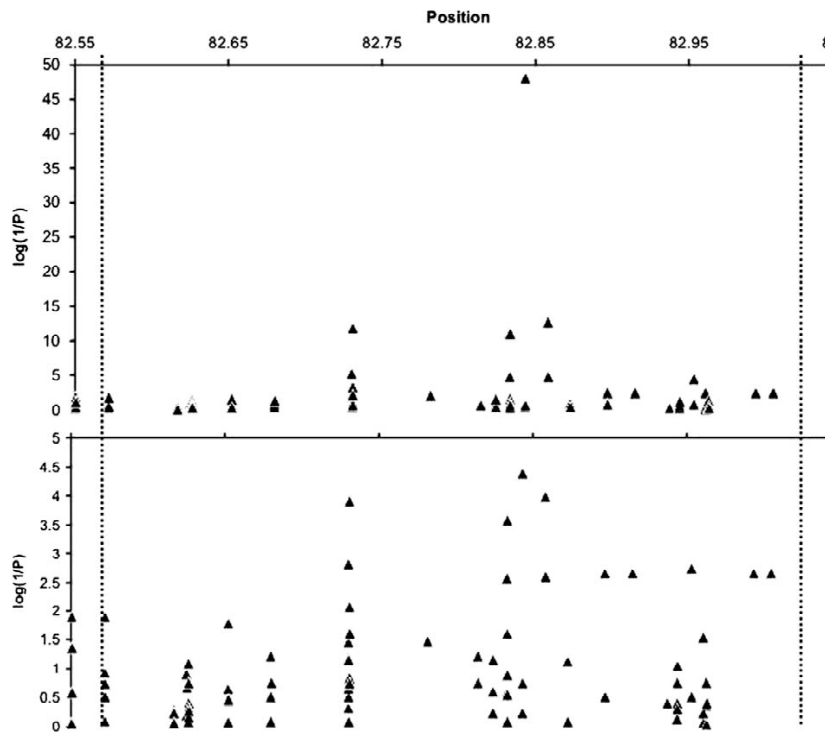
FIGURE 3.4. *In silico mapping.*
The Figure shows the association between the reproductive performance of males from the 17 inbred strains analyzed in Figure 3 and 167 diallelic variants distributed across the candidate interval defined by progeny testing. Variants are shown as triangles and the position along the horizontal axis refers to the distance in Mb from the centromere of chromosome 11. The vertical axes show the statistical significance of the association results while the dashed vertical lines denote the proximal and distal boundaries of the candidate interval. In some cases multiple variants with the same degree of association appear as a single triangle. The top plot shows the strength of the associations between genotype and litter size using the F-distribution to assess the statistical significance of the nested mixed models. The bottom plot shows the strength of the association between genotype and litter size using permutation tests to assess the statistical significance of our findings. The empirical significance estimates for the SNPs on the figure with the four most significant findings were determined exactly.

TABLE 3.2. SNPs associated with the reproductive performance phenotype

Strain	*D11Spn173-D11Spn129*	Paternal allele	1	2	3	4	5	6	7	8	9	10	11	12	Refs[†]
BS	*domesticus?*	I	ND	ND	ND	ND	ND	ND	ND	ND	ND	**ND**	**ND**	ND	[61]
NC	*domesticus?*	I	ND	ND	ND	ND	ND	ND	ND	ND	ND	**ND**	**ND**	ND	[52]
A	*domesticus*	I	C	T	T	A	T	T	G	G	G	**A**	**G**	A	‡
DBA/2	*domesticus*	I	C	T	T	A	T	T	G	G	G	**A**	**G**	A	‡
BALB/c	*domesticus*	I	C	T	T	A	T	T	G	G	G	**A**	**G**	A	‡
C57BL/6	*domesticus*	I	C	T	T	A	T	T	G	G	G	**A**	**G**	A	‡
129X1	*domesticus*	I	C	T	T	A	T	T	G	G	G	**A**	**G**	A	‡
PERA	*domesticus*	I	C	T	T	A	T	T	G	G	G	**A**	**G**	A	‡
RBA	*domesticus*	I	C	T	T	A	T	T	G	G	G	**A**	**G**	A	‡
CBA*	*domesticus*	I	ND	ND	ND	ND	T	T	G	G	G	**A**	**G**	A	[58]
C3H*	*domesticus*	I	ND	ND	ND	ND	T	T	G	G	G	**A**	**G**	A	[62, 63]
KK*	*domesticus*	I	ND	ND	ND	ND	A	C	A	A	A	**A**	**G**	G	[52]
LEWES	*domesticus*	I	C	T	T	A	A	C	A	A	A	**A**	**G**	G	‡
PWK*	*musculus*	C	C	T	T	A	A	C	A	A	A	**G**	**T**	G	‡‡
SKIVE	*musculus*	C	C	T	T	A	A	C	A	A	A	**G**	**T**	G	‡
WSB*	*domesticus*	C	ND	ND	ND	ND	ND	ND	ND	ND	ND	**G**	**T**	G	‡
DDK	*domesticus*	C	A	A	C	G	del	del	del	del	del	**G**	**T**	G	‡
MOLC	*musculus*	C	A	A	C	G	A	C	A	A	A	**G**	**T**	G	‡
PANCEVO	*M. spicilegus*	C	A	A	C	G	A	C	A	A	A	**G**	**T**	G	‡
CAST	*castaneus*	C	A	A	C	G	A	C	A	A	A	**G**	**T**	G	‡
TIRANO	*domesticus*	C	A	A	C	G	A	C	A	A	A	**G**	**T**	G	‡
ZALENDE	*domesticus*	C	A	A	C	G	A	C	A	A	A	**G**	**T**	G	‡
JF1	*castaneus*	C	A	A	C	G	A	C	A	A	A	**G**	**T**	G	‡
PERC	*domesticus*	C	A	A	C	G	A	C	A	A	A	**G**	**T**	G	‡
MOM	*molossinus*	C	ND	ND	ND	ND	ND	ND	ND	ND	ND	**ND**	**ND**	ND	[74]
CASP	*castaneus*	C	ND	ND	ND	ND	ND	ND	ND	ND	ND	**ND**	**ND**	ND	[74]

The table provides the strain name, phylogenetic origin of the haplotype in the candidate interval for the paternal gene defined by progeny testing and whether the allele at the paternal gene is incompatible (I) or compatible (C) with the maternal DDK factor. Compatible and incompatible strains are separated by a horizontal line. In addition, the table provides the allele present in each strain at each of the 12 SNPs that are significantly associated with the DDK syndrome phenotype. Alleles at the two SNPs that are most strongly associated with phenotype are underlined and in bold. The positions of the SNPs are as follows: 1, 82734817; 2, 82734820; 3, 82735318; 4, 82735343; 5, 82833126; 6, 82833146; 7, 82833386; 8, 82833571; 9, 82833572; 10, 82843176; 11, 82843476; 12, 82857766. The phylogenetic origin of the haplotypes in wild derived strains is shown as reported in Chapter 2). The phylogenetic origin of the haplotypes in hybrid strains and classical strains was determined using diagnostic alleles (See Chapter 6). ND, not determined; del, deletion.
* Additional strains tested to see if SNPs 10 & 11 predict the reproductive performance.
[†] References that support the assignment of the paternal allele.
‡This study. ‡‡ Jiri Forejt, personal communication).
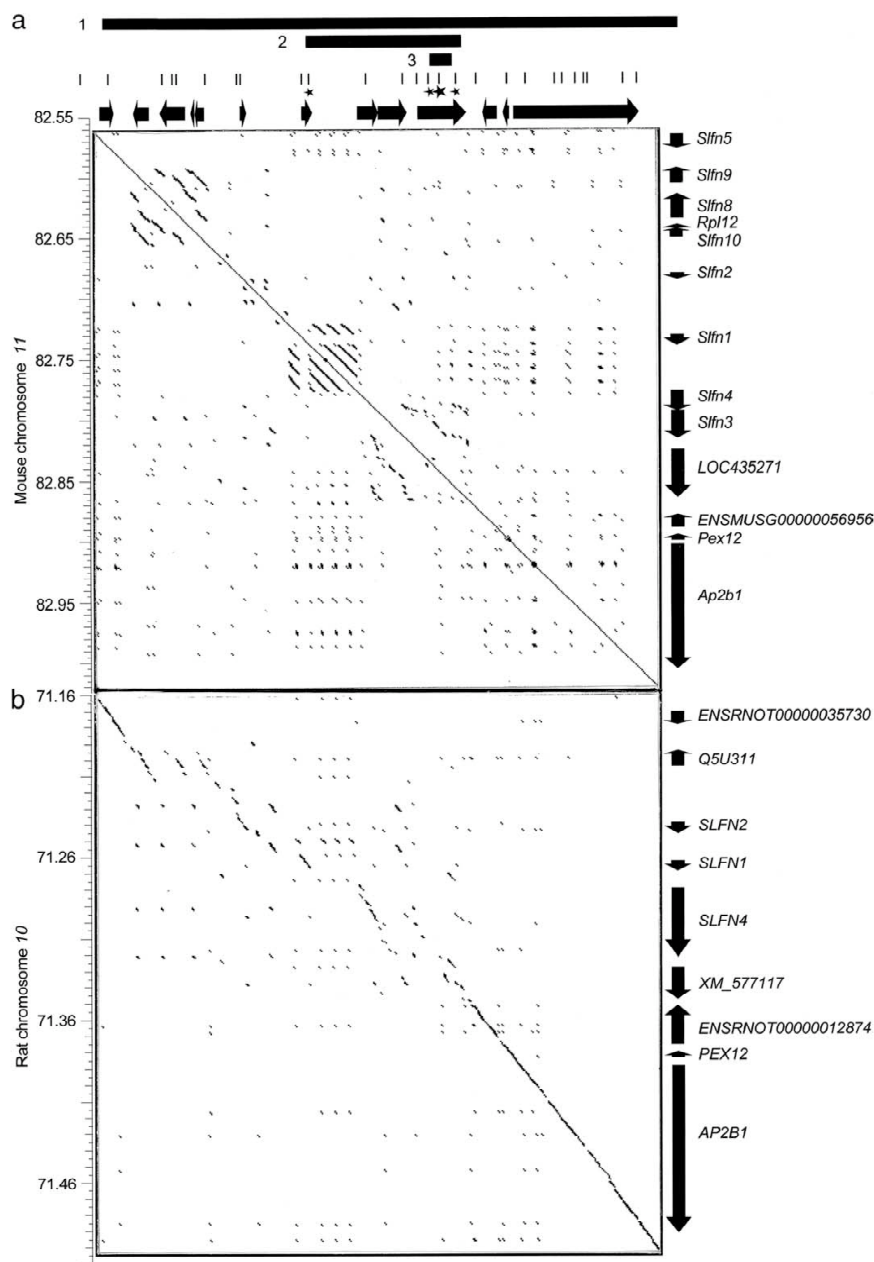
FIGURE 3.5. Rapid evolution of the *Schlafen* gene cluster in mouse. The three bars shown at the top of the figure represent the candidate intervals defined by recombinant progeny testing (1), *in silico* mapping (2) and haplotype block in complete linkage disequilibrium with the DDK syndrome phenotype (3). The 26 fragments sequenced are shown as vertical bars directly underneath. Stars denote the four fragments containing the 12 variants most strongly associated with the phenotype. The larger star indicates the fragment containing the two variants that are in complete linkage disequilibrium with the DDK syndrome phenotype. Genes are shown as black arrows. Vertical axes provide distance in Mb from the centromere in the appropriate chromosome. a) Dot plot matrix comparing the candidate interval in the mouse against itself. b) Dot plot matrix comparing the mouse candidate interval (horizontal axis) and the homologous region in rat (vertical axis). Each dot in the matrixes denotes an orthogonal 99 bp region with less than 10 mismatches. Consecutive dots form diagonal lines in regions of extended identity/similarity. For any given region on the horizontal or vertical axes the presence of multiple parallel diagonal lines denotes duplicated regions. Short scattered lines are for most part indicative of repetitive elements scattered across the region.

VI. Reducing the candidate interval for the gene encoding the maternal factor

Our analysis also provides information for the location of the gene encoding the maternal DDK factor. We mated B6 males to females carrying the chromosome that recombines between *D11Spn173* and *D11Spn178* (named "735L" in reference to the mouse in which this chromosome was initially identified). The reproductive performance observed in these crosses (Table 3.3) indicates that the "735L" chromosome carries a B6 allele at the gene encoding the maternal factor. Specifically, the average litter size in mating between homozygous "735L" / "735L" females and B6 males (7.8 ± 2.6) is inconsistent with the presence of incompatible alleles (*i.e.*, DDK) at the maternal factor. The "735L" chromosome carries DDK alleles in the proximal region and B6 alleles in the distal region and has a B6 allele at the paternal gene (Figures 3.1 and 3.2, and Table 3.1b). Therefore we can exclude the region upstream of *D11Spn173* from the candidate interval for the maternal gene.

TABLE 3.3. Reproductive performance of females carrying the critical recombination.

Om genotype			
Dam	Sire	Average litter size	SD
"735L" / DDK	B6 / B6	4.0	1.4
"735L" / B6	B6 / B6	7.0	2.9
"735L" / "735L"	B6 / B6	7.8	2.6

The table provides the genotype of dams and sires in the *Om* region and the average litter size and standard deviation for these crosses.

VII. Discussion

We have used two methods for mapping the paternal gene responsible for the polar, embryonic lethal phenotype known as the DDK syndrome. In the first approach, we identified chromosomes that were recombinant in the *Om* region and assayed the fertility of males carrying these chromosomes in crosses with females carrying the DDK maternal factor. In the second approach, we assayed the fertility of males from 17 inbred strains of

diverse phylogenetic origin in crosses with females carrying the DDK maternal factor. These experiments have resulted in a dramatic reduction in the candidate region containing the paternal DDK syndrome gene, perhaps to a single gene, as well as more general conclusions on the evolutionary history of the *Om* region and the DDK syndrome.

We have also shown that both maternal and paternal components of the DDK syndrome lie in overlapping intervals. We wish to note that the use of homozygous "735L"/"735L" females forestalls the confounding effects of modifiers of the maternal contribution to the DDK syndrome that have been reported previously, because the mode of action of these modifiers are thought to require heterozygosity at *Om* in the dam [67, 75, 77]. These data represent the first step towards the definition of a candidate interval for the maternal gene.

a) Mapping the DDK syndrome candidate region

Over the course of our investigations [55, 56, 69, 75, 90, 134], we have screened over 5,000 meioses for chromosomes 11 that were recombinant in the *Om* region. We have now determined the paternal phenotype of 18 recombinant chromosomes (Table 3.1b and Figures 3.1 and 3.2) and refined the candidate interval for the paternal gene of the DDK syndrome. Despite the substantial effort required to isolate and test these recombinant chromosomes, this traditional strategy allowed us to narrow the previous 1.5 Mb candidate interval [68, 70, 71] to a 385-465 kb interval that contains no fewer than 13 genes (Figure 3.5a). Because we appeared to have reached the effective limit of mapping resolution by traditional methods, we attempted to refine further the candidate interval by in silico association mapping of DNA sequence variants against a quantitative measure of paternal reproductive compatibility (Figure 3.3). The power of this approach depends on the number of strains analyzed, the ratio between compatible and incompatible paternal alleles among the strains and the level of linkage disequilibrium between nearby sequence variants. The

high levels of genetic diversity present in wild-derived strains (see Chapter 2) and the fact

that wild-derived *M. m. castaneus* and *M. m. molossinus* strains carry compatible alleles at

the paternal gene [74] suggested that mapping of the paternal gene might benefit from the

large number of historical recombination events captured by comparing a diverse collection

of inbred strains. Therefore, we determined the paternal phenotype of 17 strains selected to

maximize sampling across a diverse set of lineages. The males of these strains fall into one

of two non-overlapping groups with respect to the reproductive performance (Figure 3.3).

The group with the best reproductive performance includes nine strains (PANCEVO, CAST,

JF1, MOLC, DDK, SKIVE, PERC, 18 ZALENDE, and TIRANO). We conclude that these

strains carry alleles at the paternal gene that are compatible with the maternal DDK factor.

The remaining eight strains (A, 129X1, DBA/2, BALB/c, B6, PERA, LEWES and RBA) have

reduced litter sizes and are considered to be incompatible (Figure 3.3). We had sequenced

previously 7155 bp distributed across 22 fragments within the *D11Spn173 - D11Spn129*

candidate interval in each of the 17 inbred strains (see Chapter 2). Four additional

fragments were sequenced to ensure a more uniform coverage of the region (Figures 3.4

and 3.5a). We focused our analysis on the 167 variants identified in the region that are

polymorphic among strains with a M. m. domesticus haplotype based on our conclusion that

the incompatible paternal allele arose in the M. m. domesticus lineage (see below) and,

therefore, variants arising in other lineages are of little consequence for association

mapping. The 12 variants having the highest association with the reproductive performance

phenotype are located in four clustered fragments (Table 3.2, Figures 3.4 and 3.5). This

analysis suggests that the paternal gene lies within a 128 kb interval spanning four genes,

from *Slfn1* to *LOC435271* (Figure 3.5a). Phylogenetic analyses of the sequenced fragments

in the 128 kb region indicates that all incompatible strains share a common ancestor (Figure

3.6). This result is consistent with the *domesticus* origin of the incompatible allele and

indicates that the paternal gene lies within the 128 kb interval containing the distal portion of

the Schlafen gene cluster. The mouse *Schlafen* gene family is composed of 10 genes

(*Slfn1*-5, 8-10, *LOC435271* and *ENSMUSG00000056956* (Figure 3.5a)) and is thought to

arise from a common ancestor through multiple unequal recombination events [135, 136].

Previous studies classified the *Schlafen* genes into three groups on the basis of different

levels of sequence identity [135]. Schlafen proteins share a divergent AAA domain thought

to bind ATP [73]. Although the exact molecular functions of the *Schlafen* proteins are

unknown, members of this family have been implicated in regulation of lymphocyte
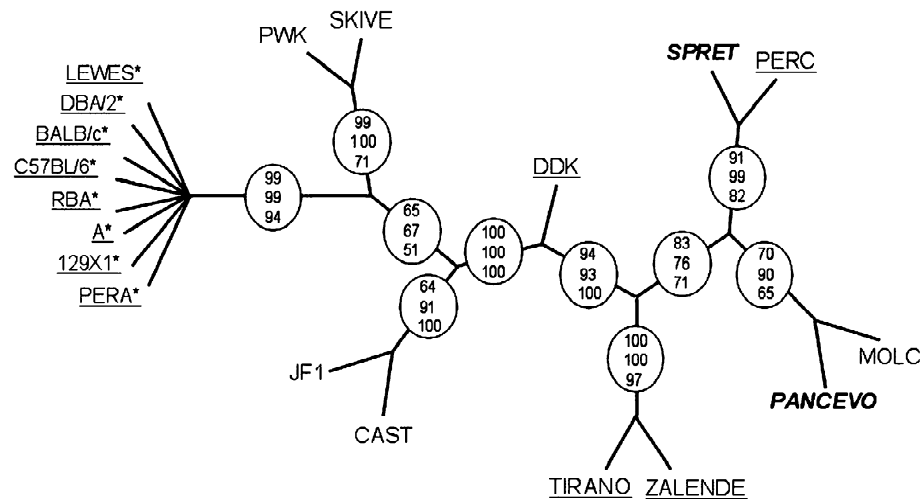
FIGURE 3.6. Phylogenetic relationships of inbred strains in the 128 kb interval defined by *in silico* mapping. The tree depicted in the figure is a consensus cladogram from three consensus trees obtained by three different phylogenetic methods (see Chapter 6). Circles denote branches that are consistent among the three trees and the numbers in the circles represent the number of times out of 100 that the branch is observed in each method; top, DNAML; middle, NEIGHBOR and bottom, DNAPARS. Underlined strains have *M. m. domesticus* haplotypes in that region. Strains in boldface and italics are derived from other species (SPRET*, M. spretus* and PANCEVO, *M. spicilegus*). Asterisks denote strains with incompatible alleles at the paternal gene. All incompatible strains are shown diverging from a single node because the internal branching order within this lineage is not consistent among trees obtained by different methods and the branching order is poorly supported within each method. The length of the branches is arbitrary.

differentiation and growth control [135, 136]. The sequence variants having the highest

significance are located in two genes, *Slfn1* and *LOC435271* (Figure 3.5a). Eight SNPs are

found in introns of the *LOC435271* gene. The four remaining SNPs are located in the coding

region of the *Slfn1* gene and result in 3 missense mutations. Given the location of these

variants (Figure 3.5a) and the complete linkage disequilibrium between the phenotype and

two SNPs in the *LOC435271* gene (SNPs 10 and 11 in Table 3.2), we propose that the

incompatible allele lies within the *LOC435271* gene, a member of the *Schlafen* gene family.

This conclusion is supported by the successful a priori prediction of the paternal phenotype

of the WSB strain on the basis of SNPs 10 and 11 (Table 3.2) as well as the successful

prediction of the two SNPs present in the C3H, KK CBA and PWK strains on the basis of

paternal phenotype (Table 3.2 and Figure 3.3). In addition, SNPs 10 and 11 define the only

haplotype that discriminate between the KK and LEWES incompatible strains and the PWK

and SKIVE compatible strains (Table 3.2). This haplotype block is in complete linkage

disequilibrium to the DDK syndrome phenotype and may span a maximum of 23.2 kb (from

variants 9 to 12 in Table 3.2).

b) Evolutionary history of the Om region and the DDK syndrome

Because the *Om* region includes two genetic factors (the maternal and the paternal

components of the DDK syndrome) with a strong effect on reproductive performance and

represents a genetic mechanism that could be linked to reproductive isolation and

speciation, the evolutionary history of this region is of interest. Phylogenetic analysis of the

sequence variants found within the 128 kb candidate interval defined by in silico mapping is

shown in Figure 3.6. Although strains from well-defined monophyletic groups, such as *M.*

musculus sp. and *M. m. domesticus* subspecies [109], are expected to cluster in a node that

does not contain strains from other clades/taxa, these expectations are not met for the

candidate region. For example, several *M. m. castaneus* and *M. m. musculus* strains

separate two groups of *M. m. domesticus* strains (DDK, TIRANO, ZALENDE and PERC in one group and LEWES, DBA/2, BALB/c, B6, RBA, A, 129X1 and PERA in the other). In the same vein, the location of the PERC and MOLC strains (two *M. musculus* strains, see Chapter 6) in the tree is incongruent with a monophyletic origin of the *M. musculus* species. Figure 3.6 demonstrates that these observations are not due to a poorly supported tree and/or to a single strain. Homoplasy can be rejected because the incongruent phylogeny is supported by multiple variants in linkage disequilibrium. Hybridization between taxa is extremely unlikely because of the lack of concordance between phylogenetic clustering and geographical origin of the wild-derived strains. For example, MOLC, a *M. m. molossinus* wild-derived strain from Japan clusters with PANCEVO, a *M. spicilegus* strain from Serbia. The most likely explanation for the observed topology is the presence of multiple ancestral variants that have segregated across clades (see Chapter 2). Although the presence of scattered ancestral variants is not surprising, per se, it is rare to find large numbers of ancestral variants that consistently contradict the expected phylogeny spanning extensive regions because normally recombination would erase the linkage disequilibrium between them. Rearrangements may suppress recombination and, therefore, in this situation sequence homology is dependent on the presence or absence of the rearrangement rather that on the overall phylogenetic relationship [137]. Rearrangements are possible candidates to explain the incongruent phylogenetic tree given the presence of multiple duplications and the fact that none of the 48 recombination uncovered between *D11Mit33* and *D11Mit35* occur in the 385 kb candidate interval defined by progeny testing (Figure 3.5a). Consistent with the idea of polymorphic rearrangements, we have also observed several kb-long insertions/deletions that are polymorphic in the set of inbred strains analyzed here (data not shown).

Rapid evolution of the Om region

        Because the candidate region for the paternal gene appears to have evolved by gene duplication, we have attempted to obtain a more complete and accurate view of the organization of the *Om* region by aligning the sequence for the candidate interval defined by progeny testing against itself (Figure 3.5a). This analysis reveals three regions that have undergone tandem duplications. Based on the level of sequence identity these duplications arose at different times. The *LOC435271* gene lies within the oldest detected duplication while the *Slfn1* gene is flanked by the newest. Further evidence of a relatively fast evolutionary rate is provided by the comparison between the candidate interval in the mouse and the homologous rat sequence (Figure 3.5b). Importantly, only the central region of the *Schlafen* gene cluster has undergone rapid evolution (shown by the lack of a well-defined diagonal line between the *Slfn1* and *LOC435271* genes in Fig. 3.5b), while the *Slfn5*, *ENSMUSG00000056956*, *Pex12* and *Ap2b1* genes have evolved at a significantly slower pace (Figure 3.5b). The former region spans the duplications and contains all the variants that are significantly associated with the reproductive performance phenotype. However, the deterioration of sequence identity in interspecies comparisons is not due to the duplications alone but is also due to enrichment for LTR and LINE repeats in the central region of the gene cluster [138].

The compatible paternal DDK syndrome allele is ancestral

        Overall, a total of 26 strains have been characterized for paternal compatibility with the maternal DDK factor (Table 3.2). Compatible and incompatible alleles are represented equally among these strains (Table 3.4). However, a striking difference in the frequency of the incompatible allele emerges when one considers the phylogenetic origin of the *D11Spn173-D11Spn178* candidate region (Test for independence between the allele at the paternal gene and phylogenetic origin of the strain, χ^2 = 11.56, 1 d.f., *p*<0.0007; Table 3.4).

In strains with a *M. m. domesticus* haplotype only 28% of the strains examined are compatible. In contrast, 100% of strains with haplotypes from other subspecies and species are compatible (Table 3.4). These results strongly support the hypothesis that the compatible allele is ancestral and the incompatible allele arose in the *M. m. domesticus* lineage after the divergence of the subspecies [74]. Based on the allele frequencies observed in wild-derived *M. m. domesticus* inbred strains, we conclude that the incompatible allele appeared in the domesticus lineage shortly after the divergence of the *M. musculus* subspecies approximately 750,000 years ago [109]. However, the incompatible allele was not fixed in the ancestors of all extant mice originating in this lineage. Furthermore, compatible and incompatible alleles are present in M. m. domesticus strains derived from

TABLE 3.4. Compatible and incompatible paternal alleles in strains with haplotypes of different phylogenetic origin in the *D11Spn173-D11Spn129* interval.

D11Spn173-D11Spn129 haplotype	Allele at the paternal gene		
	Compatible	Incompatible	Total
M. m. domesticus	5	13	18
Other species and subspecies	8	0	8
Total	13	13	26

The strains used in this analysis are listed in Table 2.

natural populations of three small and distant geographic areas. Briefly, among the three wild-derived inbred strains from Northern Italy and Switzerland, one strain, RBA, carries an incompatible allele while the other two strains, TIRANO and ZALENDE, have compatible alleles. Similarly, two strains from the Eastern US, WSB, and LEWES, have compatible and incompatible alleles, respectively. Finally, two strains from Peru, PERA and PERC, also

have discordant alleles for the DDK syndrome phenotype. Therefore, both compatible and incompatible alleles are found over a broad geographical range and have colonized the New World on more than one occasion. In contrast, only the incompatible allele is found among all classical inbred strains analyzed to date with the exception of the DDK strain. This might reflect a founder effect due to the modest number of progenitors that contributed to classical inbred strains [109]. Alternatively, selection operating in the derivation of classical inbred strains but not in the derivation of wild-derived inbred strains may be responsible for the fixation of the incompatible allele in the former. Although the incompatible allele is presumed to be neutral in the absence of the maternal DDK factor, it remains possible that this allele may have an effect on the fitness of its carriers based on the frequency and distribution of alleles in *M. m. domesticus* and classical inbred strains.

Finally, although we have been unable to separate the paternal and maternal components of the DDK syndrome through recombination, the results from Zhao and coworkers [74] and the results presented here demonstrate that they are different mutations. This conclusion is based on the type of combinations of compatible and incompatible alleles at the maternal factor and paternal gene found in inbred strains. Two types of combinations of alleles have been described previously, compatible maternal alleles and incompatible paternal alleles (found in strains such as B6, BALB/c, C3H and PERA) and incompatible maternal alleles and compatible paternal alleles (found only in DDK). In this study we report a third combination of alleles in the PERC and CAST strains, compatible maternal alleles and compatible paternal alleles (Figure 3.3). The existence of three allelic combinations demonstrates conclusively that the maternal and the paternal components are non-allelic.

CHAPTER 4. Modifiers of the DDK Syndrome

The work described in this chapter was accomplished in collaboration with Kuikwon

Kim, Dr. Daniel Pomp, Dr. Jennifer L. Moran, and Dr. David Beier. The objectives of this

study were to identify inbred strains that carry rescue modifiers of the DDK syndrome; map

major loci necessary for rescue; characterize their mode of action; and determine if

modifiers in the B6 strain map to the rescue modifier loci. I contributed significantly to the

experimental design, generation of mouse crosses, and genotyping and data analysis used

to identify and characterize modifiers. I also composed the manuscript describing these

results, which has been published in Biology of Reproduction [Ideraabdullah *et al.* 2007]

(with the exception sections V, VIb and d, and parts of section IV).

I. A sensitized screen reveals modifiers that completely rescue the DDK syndrome lethality

To identify strains carrying dominant modifiers that rescue the DDK syndrome

lethality, we screened a panel of F_1 hybrid females that were generated by crossing females

from a variety of strains to DDK males. We determined the presence of modifiers by

comparing the reproductive performance of crosses between these F_1 hybrid females and

B6 males (experimental crosses) to the reproductive performance of fully viable control

crosses. All F_1 hybrid females tested in experimental crosses carry one DDK allele at *Om*.

Therefore, in the absence of rescue, mating between F_1 females and B6 males should result

in a significant reduction (~50%) in reproductive performance in comparison to viable control

crosses [53, 54, 69, 75]. This expectation was fulfilled in crosses involving seven types of F_1

females: (CAST x DDK)F_1, (SKIVE x DDK)F_1, (129X1 x DDK)F_1, (B6 x DDK)F_1, (BALB/c x

DDK)F_1, (JF1 x DDK)F_1, (DBA/2 x DDK)F_1, and (WSB x DDK)F_1 (crosses 7-14, 17-21 in

Figure 4.1). In contrast, there were no significant differences between the reproductive

Cross	Female	No. of females	No. of litters	Ratio	p-value
1	(DDK x PERC)F_1	9	31	0.9	n. s.
2		5	13		
3	(PERC x DDK)F_1	6	21	> 1.0	n. s.
4		13	47		
5	(PERA x DDK)F_1	4	38	1	n. s.
6		2	5		
7	(CAST x DDK)F_1	8	55	0.6	< 0.0001
8		8	66		
9	(SKIVE x DDK)F_1	12	67	0.6	< 0.0001
10		6	43		
11	(129X1 x DDK)F_1	6	18	0.6	< 0.0001
12		6	22		
13	(C57BL/6 x DDK)F_1	98	362	0.4	< 0.0001
14		20	69		
15	(RBA x DDK)F_1	4	9	0.9	n. s.
16		1	3		
17	(BALB/c x DDK)F_1	27	91	≈ 0.4	< 0.0001
18		-	-		
19	(JF1 x DDK)F_1	6	63	≈ 0.5	< 0.0001
18		-	-		
20	(DBA/2 x DDK)F_1	17	41	≈ 0.5	< 0.0001
18		-	-		
21	(WSB x DDK)F_1	2	10	≈ 0.3	< 0.0001
18		-	-		

Average litter size

FIGURE 4.1. Reproductive performance of F_1 hybrid females.
Listed in columns from left to right are: the cross number, the type of F_1 female, the number of females, and the number of litters analyzed. The bar graph shows the average litter size for the corresponding cross. Filled bars represent experimental crosses. Open bars represent control crosses. Control crosses are: 2, (DDK x PERC)F_1 x DDK; 4, (B6 x PERC)F_1 x B6; 6, (PERA x DDK)F_1 x DDK; 8, (CAST x B6)F_1 x B6; 10, (B6 x SKIVE)F_1 x B6; 12, (B6 x 129X1)F_1 x B6; 14, (B6 x DDK)F_1 x DDK; 16, (RBA x DDK)F_1 x DDK). For experimental crosses 17 thru 21, we used the combined average litter size of control crosses 2 thru 14 as the expected reproductive performance (shown as cross 18 in the Figure). Standard error of the mean (SEM) bars are provided for all crosses analyzed. The first column to the right of the bar graph shows the ratio of the average litter size of the experimental cross to the average litter size of the control cross (for crosses 17 thru 21 we considered this ratio approximate because of the lack of an optimal control cross). The last column provides the significance level under the null hypothesis that the average litter size of the experimental cross is equal to that of the control cross. n.s., not significant.

performance of control and experimental crosses involving three types of F_1 females: (PERC x DDK)F_1, (PERA x DDK)F_1 and (RBA x DDK)F_1 (crosses 1-6, 15-16, Figure 4.1). In these crosses, the ratio between the average litter size in experimental and control crosses is approximately one (Figure 4.1), demonstrating extensive rescue of the lethal phenotype. We conclude that rescue is due to the presence of genetic modifiers in these three strains. Note that (RBA x DDK)F_1 females have small litter sizes in both experimental and control crosses (crosses 15 and 16, Figure 4.1). RBA carries the Rb(4.12)9Bnr Robertsonian translocation [139]. Females heterozygous for Robertsonian translocations are known to have reduced litter sizes due to the improper segregation of chromosomes during meiosis [140-143]. We conclude that the reduced litter size in the RBA crosses is unrelated to the DDK syndrome lethality. Unfortunately, the limited phenotypic range of these crosses creates difficulties in accurately determining the presence and extent of rescue. Therefore, we have only pursued the genetic analyses of the modifiers present in the PERC and PERA strains.

II. A major modifier locus maps to proximal chromosome 13

The similarity between the reproductive performance of experimental and control crosses involving (PERC x DDK)F_1 and (PERA x DDK)F_1 females is consistent with complete rescue of the lethal phenotype (*i.e.*, no embryos die from the DDK syndrome; compare cross 3 to cross 4 and cross 5 to cross 6 in Figure 4.1). Three possibilities can account for complete rescue: 1) mitochondrial inheritance; 2) a single nuclear locus acting as a maternal effect; or 3) multiple unlinked loci that are able to independently rescue the lethal phenotype. The nuclear loci may be linked or unlinked to *Om*. To discriminate between these possibilities we tested whether the rescue phenotype segregates among Om^{DDK}/Om^{B6} G_2 female offspring of experimental crosses (Figure 4.2). We analyzed the reproductive performance of crosses between B6 males and 39 Om^{DDK}/Om^{B6} G_2 females

69

with B6-DDK-PERC mixed backgrounds and 36 Om^{DDK}/Om^{B6} G_2 females with B6-DDK-

PERA mixed backgrounds (Figure 4.2a). Figure 4.2b shows the wide variation in the

reproductive performance among these G_2 females, including some that were fully viable.

Given that all of these females have identical Om^{DDK}/Om^{B6} genotypes and carry PERC or

PERA mitochondrial genomes, we conclude that the loci responsible for this variation in

phenotype cannot be mitochondrially inherited, nor can they be closely linked to Om.
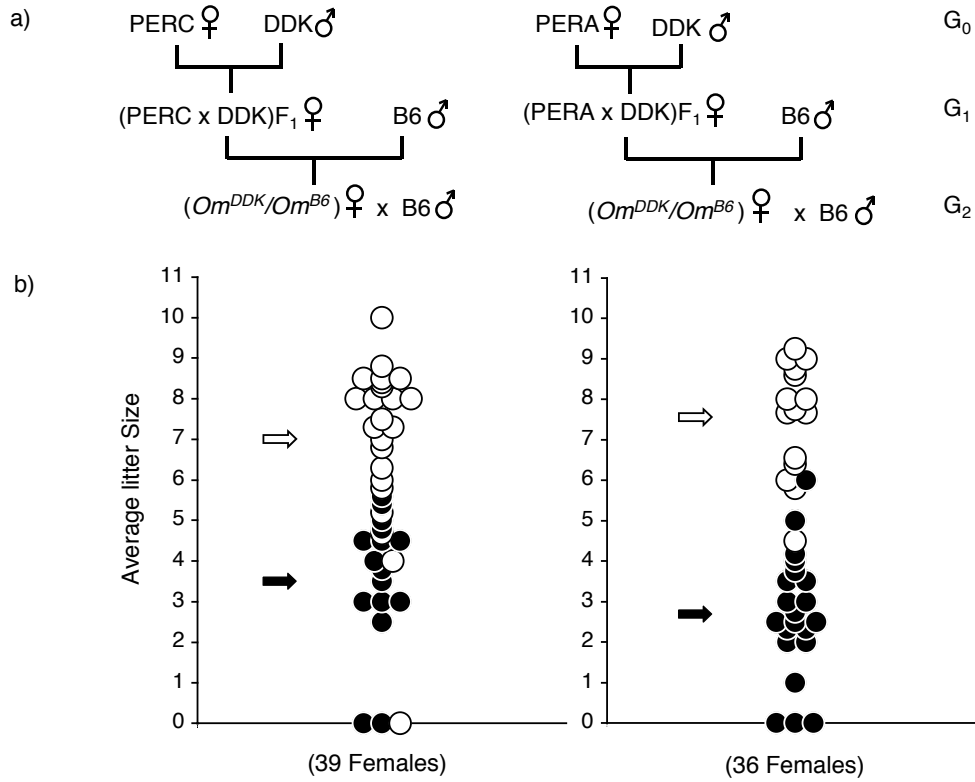
FIGURE 4.2. Reproductive performance of Om^{DDK}/Om^{B6} G_2 females.
PERC and PERA crosses are presented separately.(a) Mating scheme used to generate the
G_2 females. (b) Vertical axis represents the average litter size. Each circle represents the
reproductive performance of a single female. Open circles represent $Rmod1^{PERC}/Rmod1^{B6}$ or
$Rmod1^{PERA}/Rmod1^{B6}$ females. Filled circles represent $Rmod1^{DDK}/Rmod1^{B6}$ females. Open
arrows represent the mean of $Rmod1^{PERC}/Rmod1^{B6}$ or $Rmod1^{PERA}/Rmod1^{B6}$ females in the left
or right graph, respectively. Filled arrows represent the mean of $Rmod1^{DDK}/Rmod1^{B6}$
females. The number of females analyzed in each cross is shown in parenthesis on the
horizontal axis.

a) Rmod1

The distribution of the average litter size shown in Figure 4.2b suggests the involvement of few modifiers with major effects on rescue rather than many QTLs with small effect sizes. Therefore, we performed a whole genome scan on this panel of females to map the PERC and PERA rescue modifiers. We identified a locus on proximal chromosome 13 that is significantly linked to the variation in reproductive performance (LOD scores: 4.2 and 5.3, PERC and PERA crosses respectively, Figure 4.3). For both PERC and PERA, the maximum LOD score is observed at microsatellite marker *D13Mit135* (position 21,933,671; all base pair positions provided in this chapter are based on NCBI Build 36). As expected, rescue is associated with the presence of a PERC or PERA allele at this locus in the dam. When G_2 females are partitioned according to their genotypes at *D13Mit135*, two distinct phenotypic classes emerge. *D13Mit135 ^{DDK}/D13Mit135^{B6}* females have average litter sizes of 3.6 ± 1.7 and 2.7 ± 1.6 in PERC and PERA crosses, respectively (filled arrows in Figure 4.2b). The average litter size of individual *D13Mit135^{DDK}/D13Mit135^{B6}* females ranges from 0 to 6.0 (filled circles in Figure 4.2b). In contrast, *D13Mit135^{PERC}/D13Mit135^{B6}* and *D13Mit135^{PERA}/D13Mit135^{B6}* females have average litter sizes of 7.0 ± 2.1 and 7.5 ± 1.4, respectively (open arrows in Figure 4.2b). The average litter size of individual *D13Mit135^{PERC}/D13Mit135^{B6}* and *D13Mit135^{PERA}/D13Mit135^{B6}* females (open circles in Figure 4.2b) ranges from 0 to 10 and from 4.5 to 9.3, respectively (note that there is a single *D13Mit135^{PERC}/D13Mit135^{B6}* female with a litter size of zero, Figure 4.2b). The fact that none of the *D13Mit135^{DDK}/D13Mit135^{B6}* females have an average litter size consistent with extensive rescue indicates that a PERC and PERA allele at a locus closely linked to *D13Mit135* is necessary for complete rescue. The overlapping linkage peaks and the correlation between rescue and PERC and PERA alleles at *D13Mit135* (Figure 4.3) suggest

71
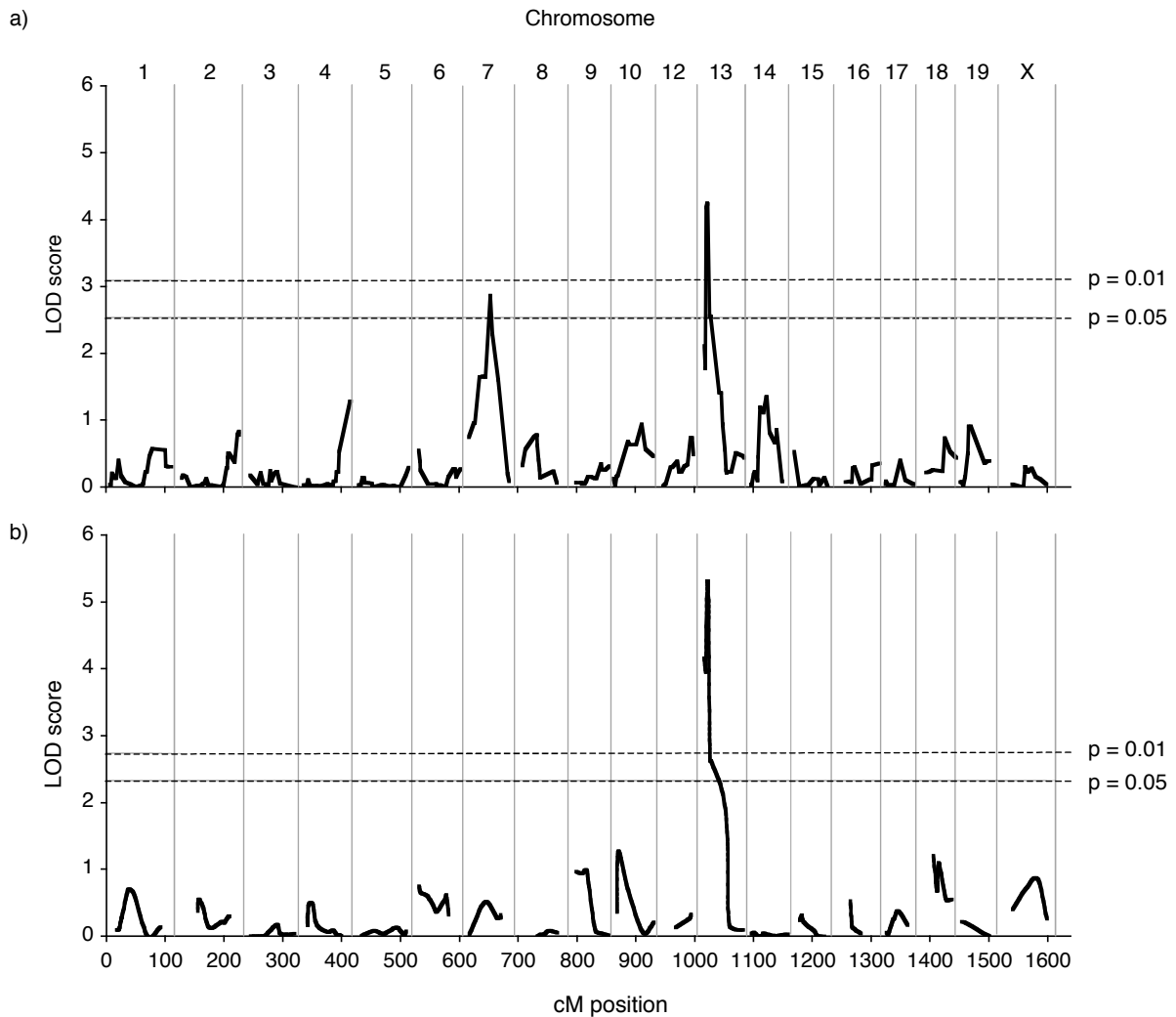
FIGURE 4.3. Linkage mapping.
a) Linkage results for G_2 PERC females. b) Linkage results for G_2 PERA females. The vertical axis represents the LOD score and the horizontal axis represents the cM position of each marker. Numbers (1-19) and letters (X), shown atop the graph, represent the respective chromosomes. The dashed horizontal line represents the significance threshold based on the *p*-value shown.

that this major modifier locus is shared by both strains. We have named this locus *Rmod1*,

for Rescue Modifier of the DDK Syndrome 1.

b) Rmod2

A second locus on chromosome 7 is also significantly associated with rescue, but only in the PERC crosses (LOD = 2.9; Figure 4.3a). The maximum LOD score is observed at SNP marker *rs13479321* (position 70,128,832). The presence of a DDK allele at this locus is associated with rescue. We have named this locus *Rmod2*, for Rescue Modifier of the DDK Syndrome 2.

III. Rescue is independent of allelic exclusion at *Om*

To test whether rescue of lethality by PERA and PERC alleles at *Rmod1* depends on allelic exclusion, we analyzed the reproductive performance of 22 Om^{DDK}/Om^{DDK} F_2 females crossed to B6 males. These females were generated by (PERC x DDK)F_1 or (PERA x DDK)F_1 intercrosses (Figure 4.4a). Figure 4.4b shows that there is extensive phenotypic variation in females with either type of genetic background (PERA or PERC). On average, these females have significantly higher reproductive performance than reported previously in lethal crosses [69, 75]. Given that these females are homozygous at *Om*, and rescue by skewed expression of a gene subject to allelic exclusion requires heterozygosity at *Om*, we must conclude that rescue of lethality by these modifiers does not require allelic exclusion.

IV. Parent of origin dependent rescue of lethality by PERA or PERC alleles at *Rmod1* and *Rmod2*

The experiments described in the previous sections demonstrate that transmission of PERC or PERA alleles at *Rmod1* or DDK alleles at *Rmod2* through the female germline leads to rescue of the DDK syndrome lethality (Figures 4.1 and 4.2). Next, we examined whether transmission of these alleles through the male germline is also able to rescue the embryonic lethality. In Chapter 3 we show that PERC males carry compatible alleles at

a) (PERC x DDK)F_1 ♀   (PERC x DDK)F_1♂
   (PERA x DDK)F_1 ♀   (PERA x DDK)F_1♂

   (Om^{DDK}/Om^{DDK}) ♀

b)
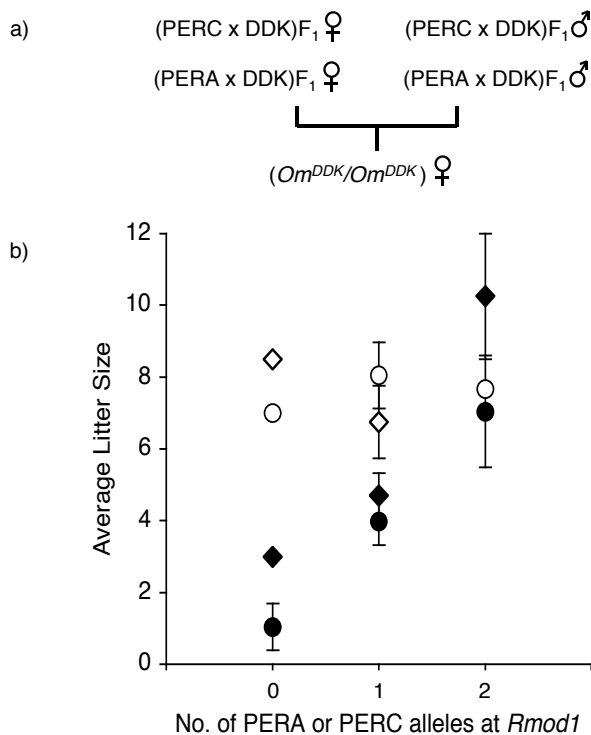
FIGURE 4.4. Reproductive performance of Om^{DDK}/Om^{DDK} F$_2$ females.
a) Mating scheme used to generate the F$_2$ females. b) The vertical axis shows the average litter size. Females were partitioned into three classes according to the number of PERC or PERA alleles they carried at *Rmod1* (horizontal axis). The number of F$_2$ females analyzed in each class is: *Rmod1^{DDK}/Rmod1^{DDK}* (PERC background), 4; and *Rmod1^{DDK}/Rmod1^{DDK}* (PERA background), 1; *Rmod1^{PERC}/Rmod1^{DDK}*, 8; *Rmod1^{PERA}/Rmod1^{DDK}*, 4; *Rmod1^{PERC}/Rmod1^{PERC}*, 3; and *Rmod1^{PERA}/Rmod1^{PERA}*, 2. Circles represent the combined average litter size of females with PERC backgrounds. Diamonds represent the combined average litter size of females with PERA backgrounds. Filled circles and diamonds represent experimental crosses (matings to B6 males) and open circles and diamonds represent control crosses (matings to DDK males). Bars denote the SEM.

the paternal *Om* locus. In contrast, we also show that PERA males carry incompatible alleles. Testing for the presence of rescue requires the sire to have at least one incompatible allele at the paternal gene. Therefore, to test transmission of alleles at *Rmod1* we generated four types of males, carrying one or two incompatible alleles at *Om* and one or two PERC or PERA alleles at *Rmod1* (cross 3,4,6, and 7, Figure 4.5). To test the transmission of alleles at *Rmod2* we analyzed three types of males, all of which carried one incompatible allele at *Om* and zero, one, or two DDK alleles at *Rmod2* (crosses 8-10, Figure 4.5). The reproductive performance of these males mated to (B6 x DDK)F$_1$ females is as predicted from their genotypes at *Om* and is not significantly affected either by the presence

74

Dam	Sire	Sire's genotype at paternal gene (*Om*)	Sire's alleles at *Rmod1*	Sire's alleles at *Rmod2*	No. of litters		% Transmission of incompatible paternal alleles at *Om*
1* (B6 x DDK) F_1	DDK	C / C	n.a.	n.d.	69		
2* (B6 x DDK) F_1	B6	I / I	n.a.	n.d.	362		
3 (B6 x DDK) F_1	(*Om* B6/B6)	I / I	*PERC/B6*	n.d.	22		
4 (B6 x DDK) F_1	PERA	I / I	*PERA/PERA*	n.d.	52		
5 (B6 x DDK) F_1	(*Om* B6/DDK)	I / C	n.a.	n.d.	33		35% (n=249)
6 (B6 x DDK) F_1	(PERA x DDK) F_1	I / C	*PERA/DDK*	n.d.	40		29% (n=296)
7 (B6 x DDK) F_1	(*Om* PERC/B6)	I / C	*PERC/PERC*	n.d.	31		34% (n=233)
8 (B6 x DDK) F_1	(*Om* B6/DDK)	I / C	n.a.	*B6/B6*	44		28% (n=309)
9 (B6 x DDK) F_1	(*Om* B6/DDK)	I / C	n.a.	*B6/DDK*	55		32% (n=479)
10 (B6 x DDK) F_1	(*Om* B6/DDK)	I / C	n.a.	*DDKDDK*	69		33% (n=413)

Average Litter Size

FIGURE 4.5. Reproductive performance of crosses between (B6 x DDK)F_1 females and males with compatible or incompatible alleles at the paternal gene at *Om*.
The figure lists the dam; the sire; the genotype of the sire at the paternal gene at *Om* (incompatible (*I*), compatible (*C*)); the sire's alleles at *Rmod1*, and *Rmod2;* the number of litters analyzed in each type of cross; and the percent of offspring inheriting the incompatible paternal allele at *Om*. The graph represents the average litter size for each adjacent cross. White bars (1,2, and 5) indicate control crosses and black bars (3,4, and 6-10) represent experimental crosses. SEM bars are provided.
n.a., Not applicable (no PERA or PERC alleles at *Rmod1*)
n.d., Genotype not determined
*Cross previously described [75].

of PERC or PERA alleles at *Rmod1,* or by the presence of DDK alleles at *Rmod2*. As shown

in Figure 4.5, experimental crosses 3 and 4 have average litter sizes consistent with

approximately 50% lethality, as expected in these types of crosses in the absence of rescue.

The average litter sizes of these two crosses are significantly different from that of viable

control cross 1 (p<0.0001) and control cross 5 (p<0.0001) (Figure 4.4). Likewise, crosses 6 -

10 have average litter sizes consistent with 25% lethality and are significantly different from

the semilethal control cross 2 (p<0.0001). The reproductive performance of crosses 6 and 8-

10 are significantly different from viable control cross 1 (p<0.05). Furthermore, although

crosses 8-10 have zero, one, or two DDK alleles at *Rmod2*, they all have similar reproductive performances (Figure 4.5), which is inconsistent with the presence of rescue associated with DDK alleles at *Rmod2*.

As a second test of embryonic lethality, we determined the transmission of paternal alleles at *Om* in the offspring of heterozygous males (Crosses 6-10, Figure 4.5). In Chapter 3 we show that embryonic lethality caused by the DDK syndrome is restricted to embryos inheriting an incompatible paternal allele at *Om*. Therefore, there should be transmission ratio distortion against incompatible paternal alleles among the surviving embryos of sires that are heterozygous for compatible and incompatible alleles at *Om*. On the other hand, if there is rescue, a reduction in the distortion levels should occur and we should observe that a greater percent of surviving embryos inherit the incompatible paternal allele at *Om* compared to control crosses. We have determined the transmission ratio in progeny of experimental crosses 6-10 (Figure 4.5). In these five crosses, we observed strong transmission ratio distortion against the incompatible paternal allele at *Om* (28-34% transmission of the incompatible paternal allele at *Om*)(Figure 4.5). These observations are not significantly different from the expected percent transmission of incompatible alleles at *Om* (~33%). In fact, when experimental crosses 6 and 7 are compared to control cross 5 the percent transmission is slightly lower. This is in complete opposition with expectations in the presence of rescue associated with PERA or PERC alleles at *Rmod1*. Furthermore, the percent transmissions of the incompatible paternal allele among crosses 8-10 are not significantly different (Figure 4.5), which is inconsistent with the presence of rescue associated with DDK alleles at *Rmod2*. Therefore, we must conclude that neither PERC nor PERA alleles at *Rmod1*, nor DDK alleles at *Rmod2*, are able to rescue lethality when transmitted through the paternal germline.

<u>V. Homozygosity for C57BL/6 alleles at *Rmod2* is associated with an increase in DDK syndrome lethality</u>

Here, we determined whether previously described recessive B6 modifiers [75-77] map to the rescue modifier loci by testing whether homozygosity for B6 alleles at *Rmod1* or *Rmod2* is associated with an increase in lethality due to the DDK syndrome. To confirm the presence of unlinked modifier loci, we compared the litter sizes of a subset of previously described Om^{B6}/Om^{DDK} females with B6-DDK mixed backgrounds to that of (B6 x DDK)F_1 females in crosses with B6 males [75]. As previously reported, these Om^{B6}/Om^{DDK} mixed background females have significantly lower reproductive performance compared to (B6 x DDK)F_1 females (p=0.0002) (Figure 4.6a). To test whether homozygosity for B6 alleles at *Rmod1* is associated with this increase in lethality, we compared the reproductive performance of $Rmod1^{B6}/Rmod1^{DDK}$ females to that of $Rmod1^{B6}/Rmod1^{B6}$ females in crosses with B6 males. The reproductive performance of these two types of crosses was not significantly different (p=0.65) (Figure 4.6b). Therefore, homozygosity for B6 alleles at *Rmod1* is not associated with an increase in lethality. To test whether homozygosity for B6 alleles at *Rmod2* is associated with an increase in lethality, we compared the reproductive performance of $Rmod2^{B6}/Rmod2^{DDK}$ females to that of $Rmod2^{B6}/Rmod2^{B6}$ females. The *Rmod2* locus was genotyped using two flanking markers, *Chr7UpSt1* (position 68,386,471), and *Chr7DwnSt4* (position 99,905,333) (Figure 4.6c & d). There is no evidence for association between homozygosity for B6 alleles at *Chr7UpSt1* and a reduction in reproductive performance (p=0.55, Figure 4.6c). However, there is evidence for association between homozygosity for B6 alleles at *Chr7DwnSt4* and a reduction in reproductive performance (p=0.006; with Bonferroni correction, significant if p<0.017, Figure 4.6d).
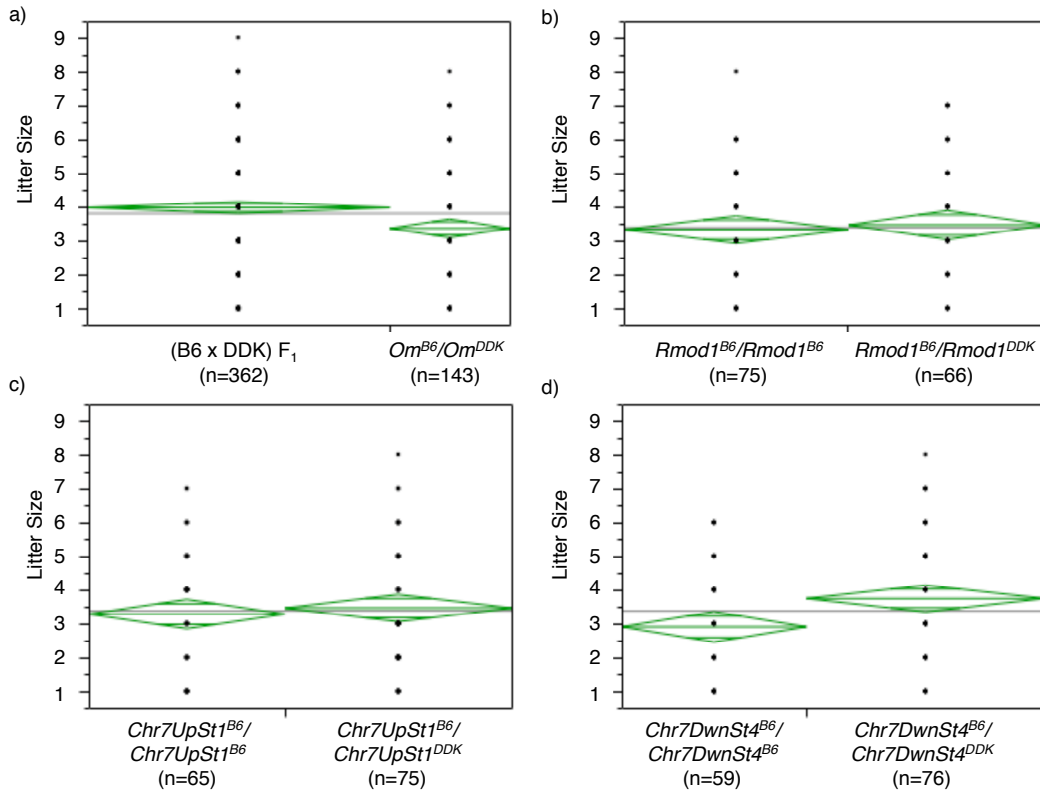
FIGURE 4.6. Association between B6 alleles at *Rmod2* and reproductive performance.
All females are Om^{B6}/Om^{DDK} , B6-DDK mixed background and were crossed to B6 males. (a)
Litter sizes from (B6 x DDK)F_1 females compared to Om^{B6}/Om^{DDK} females; (b) Litter sizes
from $Rmod1^{B6}/Rmod1^{B6}$ females compared to that of $Rmod1^{B6}/Rmod1^{DDK}$ females; (c-d) Two
loci flanking the 34 Mb candidate interval for *Rmod2* were tested, *Chr7UpSt1* (position
68,386,471); and *Chr7DwnSt4* (position 99,905,333), (c) Litter sizes from
$Chr7UpSt1^{B6}/Chr7UpSt1^{B6}$ females compared to that of $Chr7UpSt1^{B6}/Chr7UpSt1^{DDK}$ females,
and (d) Litter sizes from $Chr7DwnSt4^{B6}/Chr7DwnSt4^{B6}$ females compared to that of
$Chr7DwnSt4^{B6}/Chr7DwnSt4^{DDK}$ females. Number of litters analyzed (width of mean
diamonds) is noted in parenthesis. Graphs were created using JMP 6.0 (SAS Institute, Cary,
N.C). Height of mean diamonds (green) represent the 95% confidence intervals.

This reduction in reproductive performance is associated with the DDK syndrome lethality as

similar reproductive performances are observed in crosses between DDK males and either

Chr7DwnSt4$^{B6/DDK}$ or *Chr7DwnSt4$^{B6/B6}$* females (average litter sizes, 8.2±2.3 and 8.0±3.1,

respectively (p=0.82). It is important to note that litter sizes of zero were not included in

these analyses, therefore, we may be underestimating the association between B6 alleles and the severity of DDK syndrome lethality.

VI. Discussion

The DDK syndrome is an early embryonic lethal phenotype that has been used to study interactions between the ooplasm and the maternal and paternal genomes [50, 57, 65]. These interactions are essential to early embryonic development in mammals, a stage that is supported by ooplasmic factors. Previous studies have shown that variation in the genetic composition of oocytes can have dramatic effects in early development [12, 144-149]. Here we demonstrate the existence of modifiers that act during early development to rescue the DDK syndrome lethality and have a parent of origin effect. Data from two previous studies support the existence of rescue modifiers [53, 75]. In 1974 Wakasugi provided the first evidence of dominant rescue modifiers on the NC strain background [53]. His experiments showed that crosses between incompatible NC males and (NC x DDK)F$_1$ females are fully viable. However, the possibility of rescue modifiers in the NC strain was not discussed in that report and further investigations were not pursued. More recently, we proposed that the DDK strain carries recessive rescue modifiers [75]. However, a subsequent study failed to replicate those results [76].

a) Phylogenetic history of rescue modifiers of the DDK syndrome

Of the eleven strains we have screened for modifiers, four are *M. m. domesticus* (PERC, PERA, RBA, and WSB); one is a hybrid between *M. m. domesticus* and *M. m. musculus* (SKIVE); one is *M. m. molossinus* (JF1), one is *M. m. castaneus* (CAST); and four are classical inbred strains (B6, 129X1, BALB/c, and DBA/2). Given that *M. m. domesticus* is the subspecies that has contributed the most to the genomes of classical inbred strains [97] and NC is a classical inbred strain derived from Japanese fancy mice [104], we conclude

79

that four of the nine "*M. m. domesticus*-like" strains tested carry modifiers that rescue the

lethality in a dominant manner in crosses involving F_1 hybrid females. In contrast, none of

the three strains derived from other subspecies do. This finding suggests that the rescue

allele probably arose within the *M. m. domesticus* lineage. Our data show that these rescue

alleles are widespread within the *M. m. domesticus* lineage because they are found in

strains derived from natural populations of mice from different geographical locations (RBA

from Switzerland and PERC and PERA from Peru). However, these alleles have not been

fixed in the *M. m. domesticus* lineage (*i.e.*, WSB does not rescue, Figure 4.1). These

conclusions expand on an emerging theme in DDK syndrome research, namely, that inbred

strains derived from natural populations are valuable tools for testing, and in some cases

rejecting, longstanding hypotheses [76].

b) Defining candidate intervals for Rmod1 and Rmod2

Our linkage analyses identified two loci that are significantly associated with rescue

of the DDK syndrome: *Rmod1* on proximal chromosome 13, and *Rmod2* on chromosome 7

(Figure 4.3). We conclude that while PERA or PERC alleles at *Rmod1* are necessary for

complete rescue, they are not sufficient for complete rescue in crosses between B6 males

and Om^{DDK}/Om^{B6} G_2 females (Figure 4.2). This conclusion is based on the fact that the

combined average litter size and the range of average litter size of Om^{DDK}/Om^{B6},

$Rmod1^{PERC}/Rmod1^{B6}$ and Om^{DDK}/Om^{B6}, $Rmod1^{PERA}/Rmod1^{B6}$ G_2 females in experimental

crosses (mated to B6 males) is smaller than in control crosses (mated to DDK males). The

combined average litter size in experimental crosses is 7.2 ± 1.9 and ranges from 0 to 10;

and the combined average litter size in control crosses is 10.8 ± 0.3 and ranges from 10.5 to

11.

By considering one unit below the max LOD, the candidate intervals for the PERC

and PERA crosses overlap to define a 6 Mb candidate interval for *Rmod1* (Figure 4.7a & b).

This interval is flanked by markers *rs13481712* (position 18,146,107) and *rs13481738*

(position 29,415,412). We defined a 28 Mb candidate interval, flanked by markers

rs6160140 (position 66,160,238) and *rs13479427* (position 99,902,832), for *Rmod2* (Figure

4.7c & d).
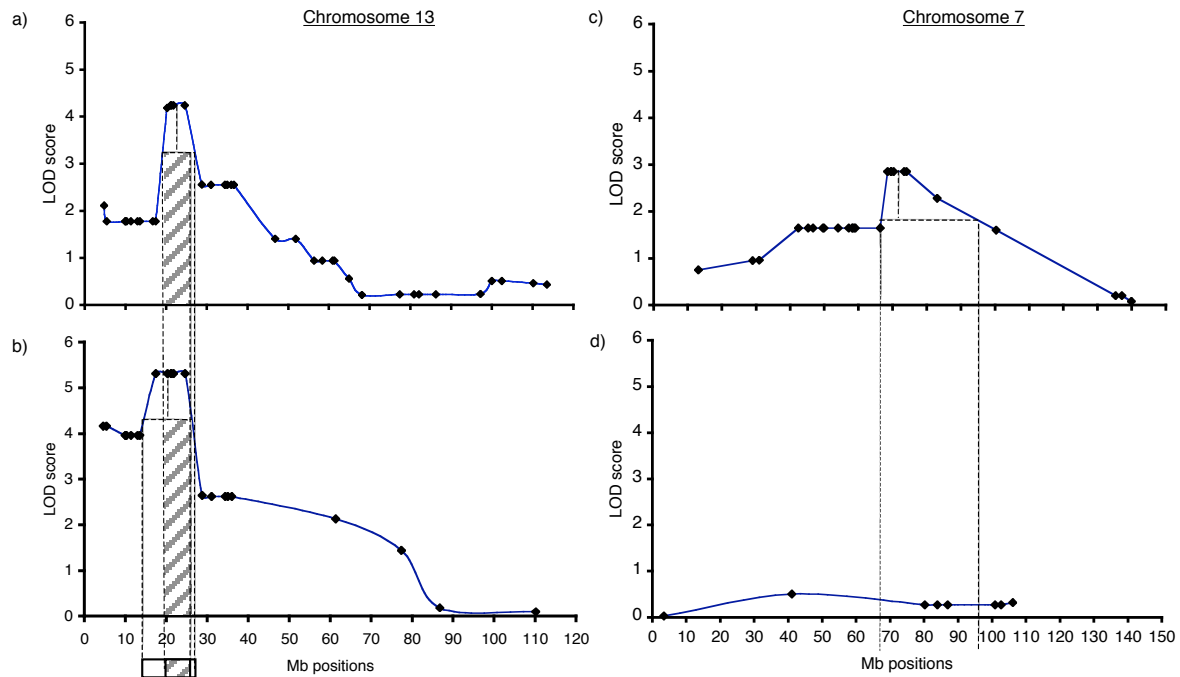
FIGURE 4.7. Candidate intervals for *Rmod1* and *Rmod2*.
Vertical axes show LOD scores and horizontal axes show Mb positions of markers analyzed
(diamond points). When one unit below the max LOD is considered, the proximal and distal
boundaries of the candidate intervals for *Rmod1* and *Rmod2* are represented by dashed
vertical lines. Graphs a & b show LOD scores for chr 13 for PERC crosses (top) and PERA
crosses (bottom), and the candidate interval for *Rmod1*. Shaded areas represent the region
of overlap of the *Rmod1* candidate intervals for PERC and PERA crosses. Graphs c & d
show LOD scores for chr 7 for PERC crosses (top) and PERA crosses (bottom) crosses,
and the candidate interval for *Rmod2*.

To reduce the size of the candidate interval for *Rmod1*, we used exclusion mapping to eliminate regions of chr 13 surrounding *D13Mit135* (marker with the max LOD score), where the genotype is not concordant with the phenotype. Because we have shown that PERA and PERC alleles are necessary for rescue, when analyzing recombinant females that have the rescue phenotype, loci with no PERA or PERC alleles were excluded from the candidate interval (Figure 4.8). We did not exclude loci by using females without the rescue phenotype because the presence of PERA or PERC alleles at *Rmod1* is not sufficient for the rescue phenotype (Figure 4.2). As shown in Figure 4.2, all females that have an average litter size greater than 6.0 also have a PERA or PERC allele at *D13Mit135*. Therefore, we only analyzed females having an average litter size of 6.5 or greater. We analyzed 13 Om^{DDK}/Om^{B6} females that carried a recombination on chr 13 within or flanking the candidate interval defined for *Rmod1* based on the LOD score (Figures 4.7 and 4.8). Seven of these females had B6-DDK-PERC mixed backgrounds and six had B6-DDK-PERA mixed backgrounds (Table 4.1). Using this analysis, the proximal boundary of *Rmod1* is defined by recombinants 3-5, and 12; while the distal boundary is defined by recombinants 8 and 10 (Figure 4.7). Therefore, we can exclude all loci upstream of *Rmod1UpSt13* and downstream of *D13Mit220* as being necessary for the rescue phenotype. This analysis defined a 5.7-6.5 Mb candidate interval for *Rmod1* (Figure 4.8). Although this method does not significantly reduce the candidate interval for *Rmod1*, it does confirm our previous results that PERA and PERC alleles at this locus, are necessary for the rescue phenotype.

Exclusion mapping was not used to reduce the 28 Mb candidate interval for *Rmod2* (Figure 4.8) due to large size of the candidate interval, the fact that alleles at *Rmod2* have a much smaller effect on the rescue phenotype, and because linkage is only observed in the PERC crosses, a limited number of recombinant females were available for the analysis.

TABLE 4.1. Crosses used to generate the recombinant females used in exclusion mapping

Cross Name	Cross Description	Genotype of recombinant females at *Om*
B1	(PERC x DDK)F$_1$ x B6	DDK/B6
B1a	((PERC x DDK)F$_1$ x B6) x B6	DDK/B6
B2	(PERA x DDK)F$_1$ x B6	DDK/B6
A2	((B6 x PERC)F$_1$ x B6) x DDK	B6/DDK
A5a	(((PERA x B6)F$_1$ x B6) x DDK) x B6	DDK/B6
A5b	(((PERA x B6)F$_1$ x B6) x DDK) x B6) x B6	DDK/B6
A6a	(((PERA x B6)F$_1$ x B6) x DDK) x B6	DDK/B6
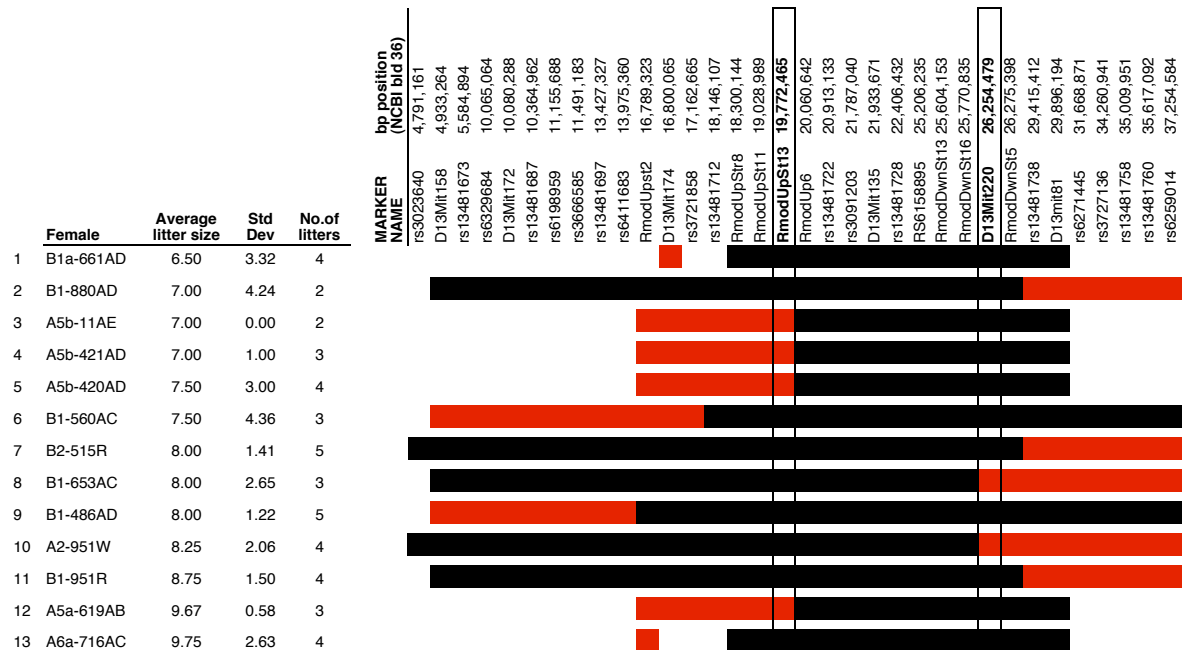
FIGURE 4.8. Exclusion mapping at *Rmod1*.
Red indicates loci that are B6/B6 or B6/DDK; Black indicates loci that are PERA/B6, PERA/DDK, PERC/B6, or PERC/DDK. Unfilled positions indicate markers for which the genotype was not determined. Markers in bold represent the proximal and distal boundaries of the *Rmod1* candidate interval as defined by this analysis.

c) Variance explained by modifier alleles at Rmod1 and Rmod2

The results of crosses involving Om^{DDK}/Om^{DDK} F_2 females confirm the critical role of the *Rmod1* locus in the rescue phenotype. The reproductive performance of Om^{DDK}/Om^{DDK} F_2 females is directly correlated to the number of PERA or PERC alleles that a female carries at *Rmod1* (range 0-2, Figure 4.4b). In fact, homozygous Om^{DDK}/Om^{DDK}, $Rmod1^{PERC}/Rmod1^{PERC}$ and Om^{DDK}/Om^{DDK}, $Rmod1^{PERA}/Rmod1^{PERA}$ F_2 females have a reproductive performance comparable to viable control crosses (Figure 4.4b). These data suggest that rescue alleles at the *Rmod1* locus may have a dosage effect on the extent of rescue from the DDK syndrome.

The presence of a DDK allele at the second modifier locus, *Rmod2,* also appears to be necessary but not sufficient for complete rescue in the PERC crosses. G_2 females with a PERC allele at *Rmod2* have an average litter size of 3.9±2.3, and G_2 females with a DDK allele at *Rmod2* have an average litter size of 6.9±1.9. *Rmod2* has no significant effect on litter size in the PERA crosses. The fact that no linkage is detected between rescue and *Rmod2* in the PERA crosses (Figure 4.3b) might be a reflection of differences in marker density between the two crosses or a consequence of the small sample size. Alternatively, it may be due to the smaller effect of *Rmod2* on the rescue phenotype or differences in the combination of alleles present in the four strains involved in our crosses.

We have tested whether there is an interaction between rescue alleles at *Rmod1* and *Rmod2.* In the PERC crosses, *Rmod1* and *Rmod2* appear to act independently of each other and to have additive effects on the rescue phenotype. For example, G_2 females with a PERC allele at *Rmod1* and a DDK allele at *Rmod2* have the best reproductive performance, 7.8±1.3 (in fact, complete rescue is never observed without this combination of alleles). The worst reproductive performance is observed in females with a DDK allele at *Rmod1* and a

84

PERC allele at *Rmod2*, 3.1±1.8. Finally, G_2 females with DDK alleles at *Rmod1* and *Rmod2* and G_2 females with PERC alleles at *Rmod1* and *Rmod2* have intermediate reproductive performances, 4.6±1.1 and 5.1±2.4, respectively.

In conclusion, we have mapped two major modifiers that explain a significant amount of the variation. *Rmod1* explains 43% and 73% of the variance in average litter size observed among of G_2 females in PERC and PERA crosses, respectively. *Rmod2* alone explains 35% of the variance in average litter size in the PERC crosses, while *Rmod1* and *Rmod2* combined in an additive model explain 60% of the variance in PERC crosses. The remaining variation may be due to other QTLs with smaller additive or epistatic effects. Given the small sample size, it is likely that such QTLs are present but were not detected. It is also possible that the remaining phenotypic variance is due to genetic and/or environmental effects on reproductive performance that are unrelated to the DDK syndrome.

d) Evidence that recessive C57BL/6 modifiers map to Rmod2

B6 and BALB/c modifiers have been previously described as being recessive additive/epistatic, unlinked to *Om* and possibly involved in allelic exclusion of the gene encoding the maternal factor [67, 75-77]. However, up until now, none of these modifiers have been mapped. By demonstrating that homozygosity for B6 alleles at *Rmod2* is significantly associated with a decrease in reproductive performance (Figure 4.5), we provide the first information about the genomic location of these modifiers. The data presented here provides evidence that recessive B6 modifiers map to *Rmod2*. Further investigation is necessary to determine whether the action of these modifiers is dependent on the presence of heterozygosity at *Om*, thereby determining their involvement in allelic exclusion of the gene encoding the maternal factor.

e) Possible modes of rescue of the DDK syndrome by PERA or PERC alleles at Rmod1

There is indirect, but compelling, evidence that the gene encoding the maternal factor undergoes allelic exclusion in the oocyte [55, 67, 75, 77]. Based on the results of our crosses between Om^{DDK}/Om^{DDK} F$_2$ females and B6 males, we conclude that rescue by PERC or PERA alleles at *Rmod1* does not require allelic exclusion at the gene encoding the maternal factor (Figure 4.4). We have also shown that rescue of lethality requires the presence of PERC or PERA alleles at *Rmod1* in the maternal germline. In contrast, transmission through the paternal germline has no discernable effect. This extreme gender dichotomy must reflect differences in expression patterns of maternally and paternally inherited alleles at modifier loci, including *Rmod1*. The interaction between the maternal DDK factor and the incompatible paternal allele has been shown to affect embryo development as early as the two-cell stage [61, 76]. If rescue requires expression of the PERC and PERA alleles at *Rmod1* prior to this lethal interaction, then a paternally inherited allele may be expressed too late to be effective. Likewise, if *Rmod1* is a gene that is only expressed in the female germline, or is subject to genomic imprinting, a paternally inherited allele would be unable to rescue. It remains to be determined whether PERC or PERA alleles at *Rmod1* rescue lethality by inhibiting the incompatible interaction between the maternal DDK factor and the incompatible paternal gene or by providing a factor necessary for normal embryonic development that is absent in incompatible crosses.

CHAPTER 5. Summary and Future Directions

<u>I. Significance of the level and structure of genetic diversity in the mouse</u>

In Chapter 2, we report our findings on the high levels of genetic diversity among wild-derived mouse inbred strains. Previous studies examining genome wide genetic diversity in laboratory inbred strains reported comparatively low levels of genetic diversity including large regions of low complexity haplotype blocks spanning two-thirds of the genome [97, 100, 101]. It was acknowledged that higher levels of genetic diversity are present among wild-derived strains, and the high levels of diversity reported for the remaining third of the genome was attributed to the presence of intersubspecific diversity resulting from the mosaic nature of the laboratory inbred strains [97]. However, the experimental design of these studies did not enable them to specifically ascertain the level of intrasubspecific variation [97, 100, 101]. In our comparisons of wild-derived *M. m. domesticus* inbred strains, we show that there is a considerable amount of intrasubspecific variation present (Table 2.1), which is in some cases comparable to the observed levels of intersubspecific variation (Figure 2.2). This finding was corroborated in more detailed investigations done recently [150, 151], one of which goes further to show that transitions from low to high variation does not correspond with the transition from intrasubspecific to intersubspecific variation [151]. Altogether, these findings demonstrate that for many of the regions of the genome reported to harbor high levels of diversity among laboratory inbred strains, the high levels of diversity observed is actually the result of intrasubspecific variation.

Prior claims of low levels of genetic diversity causing long stretches of haplotype sharing among laboratory inbred strains [97, 100, 101] implied that for a large portion of the genome, testing for natural variation associated with phenotypic variation in traits of interest would be difficult, if not impossible. Contemporary high-resolution studies, focusing on limited regions of the genome, reported that when the genomic sequence of multiple inbred strains are compared, a greater proportion of the genome is made up of high complexity haplotype blocks than previously determined [102, 116]. This suggests that the previously observed lack of haplotype diversity was in part due to the sparse placement of markers, the bias in SNP selection for genotyping, and comparison of a few closely related strains (classical inbred strains). Our analysis of SNPs (discovered through sequencing fragments across the genome in each of the 22 strains examined) confirms the presence of short "fragmented" [102] haplotype blocks when strains derived from *M. musculus* are compared (Figure 2.8).

A popular approach to QTL studies involves the use of association between haplotype and phenotype. Therefore, accurate representation of the haplotypes present across inbred strains is essential. Further characterization of the haplotype structure across laboratory inbred strains would probably best be carried out using complete sequence data for all strains. As a less thorough, but a more feasible option considering the cost and effort of sequencing the more than 400 inbred strains, haplotype structure could also be defined using genomewide SNP genotype data that more accurately reflects the level and distribution of diversity among inbred strains of all types. Our data demonstrates the importance of the inclusion of wild-derived strains in current studies to maximize genetic diversity. Therefore it is imperative that multiple strains of both classical and wild-derived origins be included in these types of SNP discovery analyses. Important steps are already being made to address these issues such as the resequencing of 15 inbred strains by

88

National Institute of Environment Health Sciences (NIEHS) and Perlegen Sciences [152, 153]. These data are publicly available [153] and represent an extensive survey of the genetic diversity present among 15 inbred strains. The four wild-derived strains (WSB, MOLF, CAST, and PWD), specifically chosen to represent each of four subspecies of *M. musculus*, were selected partly on the basis of our findings.

While the Perlegen resequencing data lack previous SNP selection biases, a recent study demonstrated that these data contain a bias for SNP discovery [150] (Yang H, Bell TA, Churchill GA, and Pardo-Manuel de Villena, F, personal communication). Investigators found a very high false negative rate for discovering SNPs within the Perlegen data (62%), which was shown to be inversely proportional to the frequency of the minor allele at a particular SNP [150] (Yang H, Bell TA, Churchill GA, and Pardo-Manuel de Villena, F, personal communication). This shows that the level of diversity present in the Perlegen dataset is an underestimate of the true levels of diversity present among the 16 inbred strains (including B6) and is skewed in comparisons involving strains that harbor the minor allele.

The high levels of diversity we observed among inbred strains was confirmed by Yang and coworkers [150] (Yang H, Bell TA, Churchill GA, and Pardo-Manuel de Villena, F, personal communication). Interestingly, they also report the presence of intersubspecific introgression among wild-derived inbred strains, which are commonly thought to represent uncontaminated inbred lines representing individual subspecies [150] (Yang H, Bell TA, Churchill GA, and Pardo-Manuel de Villena, F, personal communication). The presence of intersubspecific introgression in the genomes of wild-derived mice may explain the high number of variants we defined as "ancestral" in our study (See Chapter 2). As we discussed in Chapter 2, this type of variance may significantly inhibit our ability to accurately construct haplotype phylogeny among laboratory inbred mice. To this end, it is necessary to include

multiple wild-derived strains representing individual subspecies when determining the haplotype origin of classical inbred strains. Despite this difficulty, Yang and coworkers [150] (Yang H, Bell TA, Churchill GA, and Pardo-Manuel de Villena, F, personal communication) used the Perlegen SNP dataset and analyzed the regions of the genome that are not introgressed in the 4 wild-derived strains (~72%) to determine the subspecific origin of the genomes of the 12 classical inbred strains. They show that on average 92% of the genomes of classical inbred mice are of *M. m. domesticus* origin. This data is available through the Center for Genome Dynamics, which is an initiative in Systems Biology aimed at developing a detailed map of genetic interactions using allelic diversity, functional categories, gene expression, recombination hotspots and phenotype associations [154].

The high levels of genetic diversity available in intersubspecific crosses of mice is a very useful tool in mapping experiments, especially those set up to identify genes involved in human complex traits such as cancer, heart disease and diabetes. Not only will these types of crosses yield mice with recombinations that are necessary for high resolution mapping, these crosses will also maximize phenotypic variation, which will be useful in identifying novel gene functions and networks of interacting genes. Use of natural variation in the mouse to characterize the underlying genetic components of human disease may even be a more useful resource in some types of studies than use of artificial perturbations of the mouse genome such as mutagenesis or transgenics. The Collaborative Cross project, set up by the Complex Trait Consortium, aims to create such a resource by generating a panel of 1000 recombinant inbred (RI) lines that will capture high levels of natural variation [155]. This panel of RI lines is being generated from eight parental strains, three of which are wild-derived strains (WSB, CAST, and PWK) that were specifically included to maximize the level of genetic diversity, as exemplified by our study. These lines are expected to be available to the public within the next 5 years [155].

Studies on the DDK syndrome began over 45 years ago. Initially, interest in the DDK syndrome focused on its potential as a model to study early embryonic defects and possibly speciation. However, with the discovery of genomic imprinting in the 1980s, there was much interest in the parent of origin dependent aspect of the DDK syndrome. Later, the meiotic drive system, found to be linked to the locus responsible for the DDK syndrome, sparked renewed interest in the DDK syndrome and its implications in evolutionary biology. Despite the numerous intriguing characteristics of the DDK syndrome, and its proposed relevance in developmental research, the popularity of DDK syndrome research has declined over the years. This is most likely due, in part, to the complexity of the phenotype, the uniqueness of the phenotype (in the sense that it is restricted to crosses involving the DDK strain), and a perception that DDK syndrome studies lack significant relevance to human disease.

The work covered in this dissertation represents major breakthroughs in the study of the DDK syndrome and sheds new light on its usefulness as a genetic model. These results were achieved by utilizing a combination of approaches in a relatively new fashion and required the use of the mouse genome sequence and annotation, high-throughput SNP genotyping panels, and a variety of web-based bioinformatics databases and computational programs (See Chapter 6). Most importantly, my research would not have been possible without the inclusion of wild-derived inbred strains, which greatly increased our ability to map and characterize genetic components of the DDK syndrome.

a) The genetic architecture of the DDK syndrome

Prior to this study, the genetic makeup of the DDK syndrome consisted of two major components: the gene encoding the maternal factor, at which the incompatible allele is only carried by the DDK strain; and the paternal gene, for which seven classical inbred strains

were shown to carry the incompatible allele and only three strains (DDK, MOM, and CASP) were shown to carry the compatible allele (Figure 5.1). The paternal gene was mapped to a 1.5 Mb interval on chr 11 [70, 71], to which the gene encoding the maternal factor was later shown to be linked [69]. However, a candidate interval was not defined for the gene encoding the maternal factor, and it remained unknown as to whether the two components were allelic. Previous studies also demonstrated the presence of multiple recessive

Strains	Maternal Factor	Paternal Gene	Rmod1	Rmod2	References
A	Compatible	Incompatible	n.d.	n.d.	†
DBA/2	Compatible	Incompatible	No Rescue	n.d.	†
BALB/c*	Compatible	Incompatible	No Rescue	Increased Lethality ?	[67], †
B6	Compatible	Incompatible	No Rescue	Increased Lethality	[51], †
129X1	Compatible	Incompatible	No Rescue	n.d.	†
C3H	Compatible	Incompatible	n.d.	n.d.	[62], [63], †
CBA	Compatible	Incompatible	n.d.	n.d.	[58]
KK	Compatible	Incompatible	n.d.	n.d.	[51], [52]
NC*	Compatible	Incompatible	Rescue?	Rescue?	[51], [52], [53]
BS	Compatible	Incompatible	n.d.	n.d.	[61]
DDK	Incompatible	Compatible	No Rescue	Rescue	[51], †
RBA*	Compatible	Incompatible	Rescue?	Rescue?	†
LEWES	Compatible	Incompatible	n.d.	n.d.	†
PERA	Compatible	Incompatible	Rescue	No Rescue	†
PERC	Compatible	Compatible	Rescue	No Rescue	†
WSB	Compatible	Compatible	No Rescue	n.d.	†
TIRANO	Compatible	Compatible	n.d.	n.d.	†
ZALENDE	Compatible	Compatible	n.d.	n.d.	†
SKIVE	Compatible	Compatible	No Rescue	n.d.	†, ‡
PWK	Compatible	Compatible	n.d.	n.d.	†
JF1	Compatible	Compatible	No Rescue	n.d.	†
MOLC	Compatible	Compatible	n.d.	n.d.	†
MOM	Compatible	Compatible	No Rescue	n.d.	[74]
CASP	Compatible	Compatible	No Rescue	n.d.	[74]
CAST	Compatible	Compatible	No Rescue	n.d.	†
PANCEVO	Compatible	Compatible	n.d.	n.d.	†
	465-632 kb (chr 11)	25 kb (chr 11)	5.7-6.5 Mb (chr 13)	28 Mb (chr 7)	

FIGURE 5.1. The genetic architecture of the DDK syndrome. Phylogeny of *Mus* species and *M. musculus* subspecies based on description by Guenet and Bonhomme [109]. Subspecific origin of strains is assigned based on [156] and our studies (see Chapter 2). The size and chromosomal location of the candidate interval for each locus is listed below the chart. n.d., not determined.
*These strains have been shown to carry modifiers, but it has not been determined whether or not the loci responsible map to *Rmod1 or Rmod2*.
† This study
‡ Jiri Forejt (personal communication)

modifiers of the DDK syndrome in B6 and BALB/c that are unlinked to *Om* and cause an increase in the number of pups affected by the DDK syndrome [67, 75, 77]. However, the genomic locations of these modifier loci were unknown.

Here, we have characterized the DDK syndrome compatibility of 16 additional strains; significantly reduced the interval for the paternal gene to a 23 kb interval that encompasses a single gene; defined a candidate interval for the gene encoding the maternal factor; and last but not least, mapped additional loci involved in the DDK syndrome (Figure 5.1). These include two modifier loci, *Rmod1* and *Rmod2*, at which PERA and PERC alleles, and DDK alleles, respectively, are responsible for complete rescue of the DDK syndrome phenotype. In addition, more than seven years after recessive B6 and BALB/c modifiers were shown to increase DDK syndrome lethality, here we show that homozygosity for B6 alleles at *Rmod2* is significantly associated with an increase in lethality. Our observation of the three possible combinations of alleles for the gene encoding the maternal factor and the paternal gene (incompatible–compatible, compatible–incompatible, and compatible–compatible; Figure 5.1) demonstrates that alleles at these two loci can segregate independently of one another and thus, confirms that these two components of the DDK syndrome are non-allelic. However, it is possible that the causative mutations may reside in the same gene.

As illustrated in Figure 5.1, this study has greatly expanded our previous understanding of the genetic architecture of the DDK syndrome. Among the 26 strains tested to date, DDK is the only strain carrying the incompatible allele for the maternal factor. On the other hand, all 10 classical strains (excluding DDK) tested to date, carry the incompatible allele of the paternal gene, while alleles at the paternal gene segregate among wild derived strains of *M. m. domesticus* (Figure 5.1). Furthermore, rescue alleles at *Rmod1* and *Rmod2* are only present among *M. m. domesticus* strains (Figure 5.1). Although, data

from previous studies suggests the presence of rescue modifiers in NC, there is no evidence that these modifiers map to *Rmod1* or *Rmod2*. Taken together, these data suggest that the alleles involved in the incompatibility of parental genomes underlying the DDK syndrome, arose in the *M. m. domesticus* lineage prior to the generation of laboratory inbred strains. Further analysis of strains of other subspecific/specific origin is necessary to confirm this hypothesis. Finally, the presence of modifiers of the DDK syndrome in six strains (B6, BALB/c, NC, PERA, PERC, and RBA), of which the latter three are not closely related to the DDK strain, lends support to the proposal that the molecular mechanisms underlying the DDK syndrome are not just an oddity of the DDK strain but may indeed have an important function in normal mammalian embryonic development.

b) The presence of an allelic series at Om

In all studies to date, investigators have assumed that only two functionally different alleles are found at the gene encoding the maternal factor at *Om*: one allele that encodes the incompatible maternal factor, and another allele that is compatible (alleles *om* and *OM* respectively, see Chapter 1). My preliminary data suggests that there are functional differences between compatible alleles at the gene encoding the maternal factor. While crosses involving heterozygous Om^{B6}/Om^{DDK} females are semilethal, crosses involving heterozygous Om^{PERA}/Om^{DDK} or Om^{PERC}/Om^{DDK} females exhibit a partial rescue phenotype (*i.e.* fewer embryos die). This effect is observed in the absence of rescue modifiers at *Rmod1*. Figure 5.2 shows that in crosses with B6 males, Om^{PERA}/Om^{DDK}, $Rmod1^{B6}/Rmod1^{DDK}$ and Om^{PERC}/Om^{DDK}, $Rmod1^{B6}/Rmod1^{DDK}$ females with B6-DDK-PERA or B6-DDK-PERC mixed background have significantly higher reproductive performance compared to Om^{B6}/Om^{DDK}, $Rmod1^{B6}/Rmod1^{DDK}$ females (p=0.04). It is important to note that all of these females carry a single DDK allele at *Rmod2* thereby ruling out a possible effect of DDK

alleles at *Rmod2* on this rescue phenotype. These data provide evidence that the presence

of PERA or PERC alleles at the gene encoding the maternal factor at *Om*, or at loci linked to

Om, is able to modify the severity of the DDK syndrome. In fact, the rescue effect of PERA

or PERC alleles at *Om* may be necessary for the complete rescue of lethality observed in

crosses between (PERA x DDK)F_1 and (PERC x DDK)F_1 females and B6 males (Chapter 4,

Figure 4.1). As this level of rescue was not observed in crosses between Om^{B6}/Om^{DDK} G_2

females and B6 males (Chapter 4, Figure 4.2).
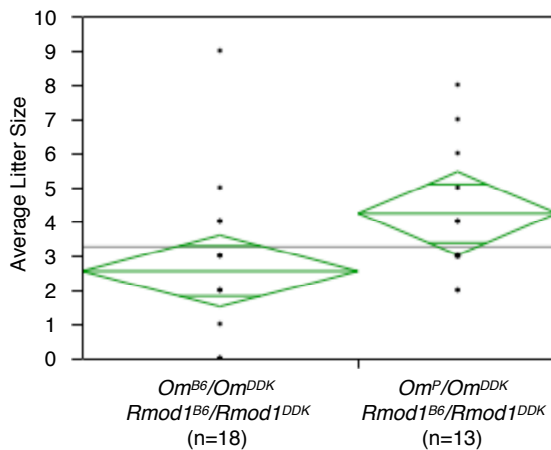
FIGURE 5.2. The presence of PERA or PERC alleles at *Om* has a rescue effect. Height of mean diamonds (green) represent the 95% confidence intervals. Reproductive performances are combined for 4 Om^{PERA}/Om^{DDK} $Rmod1^{B6}/Rmod1^{DDK}$ and 9 Om^{PERC}/Om^{DDK} $Rmod1^{B6}/Rmod1^{DDK}$ females (Om^P/Om^{DDK} $Rmod1^{B6}/Rmod1^{DDK}$). Number of females analyzed (width of mean diamonds) is noted in parenthesis. Graph was created using JMP 6.0 (SAS Institute, Cary, N.C).

Our finding of functional differences between compatible alleles at the gene encoding

the maternal factor at *Om*, may also have some bearing on the allelic exclusion hypothesis

illustrated in Figure 1.4. Under this hypothesis, the gene encoding the maternal factor is

subject to allelic exclusion in the oocyte. Allelic exclusion is thought to be the cause of the

semilethal nature of the cross and the observed 50:50 (DDK : non-DDK) segregation of

alleles at *Om* in offspring of heterozygous females (Table 1.2). It was proposed that the

increase or decrease in DDK syndrome lethality associated with B6 and BALB/c or DDK

modifiers (respectively) is the result of skewing of allelic exclusion of the gene encoding the

maternal factor at *Om* [55, 67, 75-77]. Although we have demonstrated that heterozygosity

at *Om* is not necessary for the rescue phenotype associated with PERA and PERC alleles at *Rmod1*, these results do not rule out the possibility that skewing of allelic exclusion at *Om* may also contribute to rescue. Hence, the presence of PERA or PERC alleles at *Om* may reduce the proportion of ooctyes expressing the maternal DDK factor. Without knowing which gene encodes the maternal factor, we are unable to directly test this hypothesis by testing for differential expression of the gene encoding the maternal factor.

These results are the first demonstration of an allelic series at the *Om* locus, a fact that could not only alter the way we approach studies of the maternal factor of the DDK syndrome, but may also have implications in our understanding of linked loci such as the paternal gene of the DDK syndrome as well as loci responsible for meiotic drive at *Om*. To confirm that the difference in reproductive performance is linked to PERA or PERC alleles at *Om*, we will test congenic Om^{B6}/Om^{DDK}, Om^{PERA}/Om^{DDK}, Om^{PERC}/Om^{DDK} females that are B6-DDK across the rest of the genome, as a means of excluding the possibility that PERA or PERC alleles at other loci unlinked to *Om* are responsible for the rescue effect.

III. Future directions for DDK syndrome studies

a) Recessive modifiers and allelic exclusion at Om

We have shown that homozygosity for B6 alleles at *Rmod2* results in an increase in lethality. Next, we should test whether this effect is dependent on allelic exclusion at *Om*, by determining whether or not it occurs in crosses involving females that are heterozygous at the *Om* locus and homozygous for B6 alleles at *Rmod2*. Furthermore, it would be interesting to test whether homozygosity for BALB/c alleles at *Rmod2* has a similar effect. B6 and BALB/c have the same subspecific origin for the *Rmod2* candidate interval (*M. m. domesticus*) [150] (Yang H, Bell TA, Churchill GA, and Pardo-Manuel de Villena, F, personal communication). It is also necessary to test whether they share a unique haplotype and thus are IBD for this region.

b) Refining the candidate interval for Rmod1

        As discussed above, I was unable to significantly reduce the size of the *Rmod1*

candidate interval using exclusion mapping on a panel of 13 recombinant females. As an

alternate method, I decided to use haplotype association mapping, using a SNP dataset

generated by the Genomics Institute of the Novartis Research Foundation (GNF)(obtained

through personal communication with Tim Wiltshire) [101]. This dataset is the only set

available in which all of the strains screened for the rescue phenotype have been

genotyped. Phylogenetic analysis using the GNF dataset reveals that PERA and PERC

consistently share a haplotype that is distinct from the haplotype of the other 9 strains (B6,

BALB/c, DBA/2, 129x1, DDK, CAST, SKIVE, JF1 and WSB), for the 6.5 Mb interval for

Rmod1 (Figure 5.3). I also analyzed the phylogenetic relationships between the strains for

two additional 1 Mb long regions of chr 13, one of which is proximal to *Rmod1* (positions

5,000,000-6,000,000), and one of which is distal to *Rmod1* (positions 78,000,000-

79,000,000). PERA and PERC do not share a unique haplotype in either of these two

regions of chr 13 (data not shown). This supports our hypothesis that the allele at *Rmod1*

that is necessary for rescue is IBD in PERA and PERC. I did not observe any SNPs in the

GNF dataset for the *Rmod1* interval that were in complete linkage disequilibrium with the

presence or absence of rescue. This is most likely due to the fact that the SNPs in the GNF

dataset were discovered in a small set of mostly classical inbred strains (B6, 129SvIm, C3H,

DBA/2, A, BALB/c, CAST and SPRET) [101]. Therefore, there is an excess of SNPs in this

dataset that are polymorphic between the classical strains (all of the classical strains

screened do not have the rescue phenotype, Chapter 4, Figure 4.1) and the likelihood of

finding a SNP which is not polymorphic between the classical strains is very low.

FIGURE 5.3. Phylogenetic relationships between inbred strains within the 6.5 Mb candidate interval for *Rmod1*. Strains that have the rescue phenotype are underlined. Wild-derived strains are in bold. were in complete linkage disequilibrium with our phenotype.
a) The tree is a consensus cladogram from three consensus trees obtained by three different phylogenetic methods (see Chapter 6). Circles denote branches that are consistent among the three trees and the numbers in the circles represent the number of times out of 100 that the branch is observed in each method; top, DNAPARS; middle, DNAML, and bottom, NEIGHBOR. b) This phylogram was drawn using maximum parsimony (DNAPARS) to show branch lengths.

To increase our power for haplotype association mapping to reduce the candidate interval, it would be very useful to identify additional strains that carry rescue modifiers at *Rmod1* or *Rmod2*. It would also be beneficial to use a SNP dataset generated using SNPs discovered in a more genetically diverse panel of strains or SNPs discovered through sequencing the strains of interest.

An alternate method of reducing the candidate interval for *Rmod1* would be to perform exclusion mapping on a larger panel of PERA-B6-DDK or PERC-B6-DDK mixed background recombinant females (see Chapter 4). Generating these females through alternating backcrosses to B6 or DDK may be necessary to avoid the effects of recessive B6 or DDK modifiers.

c) Candidate genes within the Rmod1 interval

Our criteria for identifying candidate genes within the 6.5 Mb candidate interval for *Rmod1* include: a) expression in the female germline; b) known or predicted function in early embryonic development (*i.e.* cell growth, cleavage or differentiation, cell-cell communication, or trophoblast development); and c) evidence of interactions between genes at *Rmod1* and genes at *Om*, or evidence that these genes are in the same molecular pathway.

The candidate interval for *Rmod1* contains 134 known genes, only 13 of which are expressed in the oocyte, (Figure 5.4). Based on Ensembl gene ontology (GO) descriptions [73], five of these genes are transcription regulators (*Trim27*, *Zfp184*, *Zfp192*, *Abt1*, and *Ttrap*), two of these genes are involved in oxidative metabolism (*Cox5b* and *Cmah*), one gene is involved in cell motility (*Elmo1*), one gene is involved in chromosome organization (*Hist1hbc*), one gene is involved in cell cycle regulation (*Gmnn*), one gene is involved in glycolysis (*Lrrc16*), and two gene functions are unknown (*Hbld2* and *Them2*) (Figure 5.4).

The known functions of most of these genes are not directly relevant to any of the specific phenotypes associated with the DDK syndrome. However, a recent publication describing the phenotype of *Gmnn* knockout mice showed that *Gmnn$^{-/-}$* embryos have abnormal nuclei, defects in cell proliferation, cell adhesion and trophectoderm development, dispersed inner cell mass, lack a blastocoel cavity, are irregular in size and die between E3.5 and E7.5 [157]. All of these phenotypes, with the exception of abnormal nuclei, have been reported to occur in embryos affected by the DDK syndrome [52, 54, 61, 62]. Several
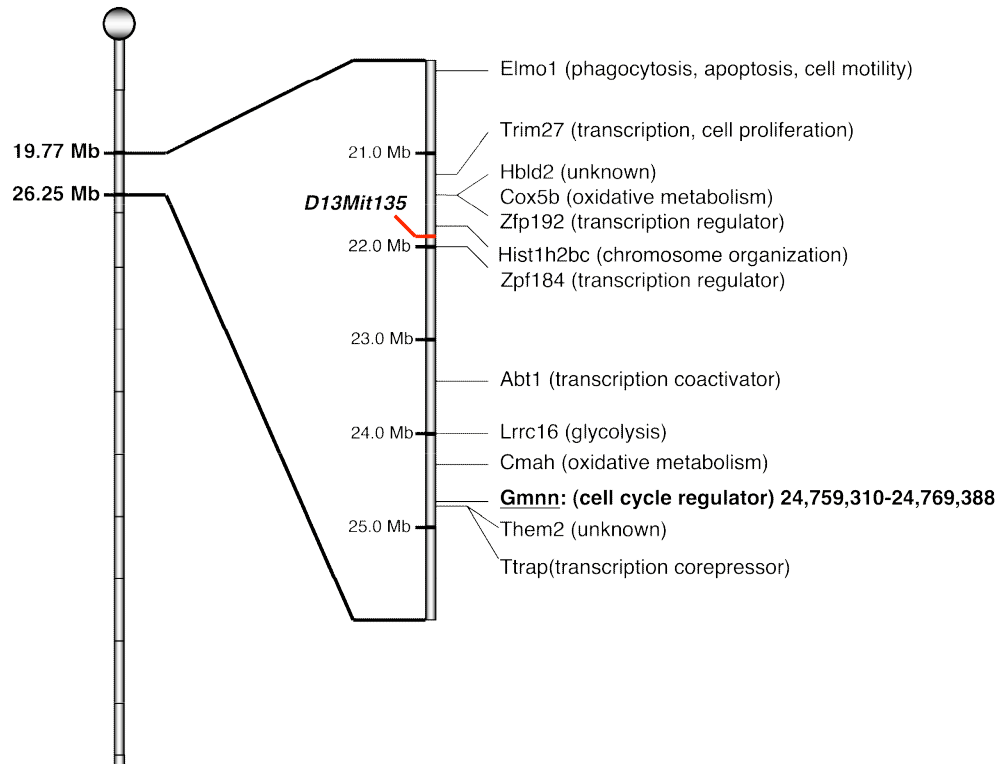
FIGURE 5.4. Genes within the *Rmod1* candidate interval that are expressed in the oocyte. Expression is based on NCBI BLAST queries and Unigene expression profiles (See Chapter 6).

publications provide evidence that *Gmnn* is involved in cell cycle regulation during meiosis, DNA replication inhibition, entry of the cell into mitosis, chromosome segregation associated with centrosome duplication, and cell differentiation [158] [159-164]. There is a significant amount of evidence that *Gmnn* inhibits *Cdt1*, *Chk1*, and *Cyclin B1*, thereby inhibiting cell division and mitosis [158, 161, 165]. This suggests that *Gmnn* may be in the same biological pathway as members of the *Schlafen* gene cluster, which have been implicated in cell cycle regulation by inhibiting cyclins *D1*, *B1*, *E2* and *A*, and cell differentiation [135, 136, 166]. Interestingly, at least five of the ten *Schlafen* genes are found in conserved synteny in humans.

100

There are no prior reports of interactions between *Gmnn* and any of the *Schlafen* genes. Further studies to determine the relationship between *Gmnn* and *Schlafen* genes may introduce novel functions in embryonic development as well as answer fundamental questions about the DDK syndrome. It would be useful to sequence *Gmnn* for the strains that have been analyzed for the rescue phenotype and determine whether any of the genetic variation in that region is concordant with the phenotype. Alternatively, we could test whether alleles at *Gmnn* are differentially expressed in the oocytes of rescuing vs. non-rescuing strains; and/or whether PERA or PERC *Gmnn* transcript is able to rescue the lethal phenotype when introduced into the ooplasm of (DDK x B6)F_1 embryos.

Overall, despite the complexity of the DDK syndrome phenotype, the results presented in this dissertation demonstrate that by using a combination of creative approaches, significant advances have been made in DDK syndrome research. We have shown that alleles at several loci involved in the phenotype are not unique to the DDK strain, but are present among natural populations of mice, thereby implicating the involvement of these loci/genes in normally occurring biological processes. Continuing studies of the DDK syndrome will likely provide answers to interesting and biologically relevant questions regarding the fundamental components of early embryonic development. In particular, the DDK syndrome is one of few models available that can be used to study the molecular mechanisms underlying maternal effects, parent of origin effects, and possibly allelic exclusion. Thus, these studies will provide much-needed insight into interactions between the parental genomes and the ooplasm that are essential to normal mammalian development.

CHAPTER 6. Materials and Methods

I. Mouse strains

      All inbred mouse strains were originally obtained from The Jackson Laboratory (Bar Harbor, ME, USA) with the exception of JF1/Ms and DDK/Pas that are maintained by Terry Magnuson and Fernando Pardo-Manuel de Villena, respectively, at the University of North Carolina at Chapel Hill (UNC-CH). The DDK/Pas strain was originally obtained from Charles Babinet of the Institute Pasteur (Paris, France). Based on their origin and known genetic make up [156], these strains are classified into two groups and assigned the following taxonomy: 1) Wild-derived strains: *Mus spretus*: SPRET/EiJ; *Mus spicilegus*: PANCEVO/EiJ; *M. m. castaneus*: CAST/EiJ, CASA/EiJ; *M. m. castaneus x M. m. domesticus*: CALB/RkJ (see Chapter 2 Discussion); *M. m. musculus*: CZECH1/EiJ and PWK/Ph; *M. m. musculus x M. m. domesticus* hybrid: SKIVE/Ei; *M. m. molossinus*: MOLC/RkJ, JF1/Ms (often described as *M. m. molossinus*, however, it is a fancy mouse with considerable contribution of *M. m. domesticus* [156]); *M. m. domesticus*: PERA/EiJ, PERC/EiJ; ZALENDE/EiJ, TIRANO/EiJ, LEWES/EiJ, RBA/DnJ and WSB/EiJ; and 2) Classical strains: BALB/cJ, C57BL/6J (B6), DBA/2J, A/J, 129X1/SvJ, C3H/HeJ and DDK/Pas. Inbred strain DNA was obtained from The Jackson Laboratory (Bar Harbor, Me, USA). All other DNA was extracted from tail biopsies as described previously [92]. All animal protocols were approved by the Institutional Animal Care and Use Committees of the University of North Carolina at Chapel Hill.

<u>II. Characterizing genetic diversity</u>

a) Primer design and PCR for sequencing

Cct6b-Ap2b1 region (Chromosome 11: positions 82,349,476-82,993,807; All base pair positions listed in Chapter 2 are based on the Ensembl v31 of the NCBI m33 mouse genome assembly [167]): We selected 39 candidate segments for analysis, which were approximately evenly spaced over the 650 kb genomic region. Sequences were selected to avoid duplicated and repeated regions in the primers, detected by BLAST and RepeatMasker software. At least one SNP was known to be present in the Celera database [168] in 24 fragments. Primers were designed to amplify an average of 450bp using 'PrimerQuest' [169].

Il9r gene (Chromosome 11: positions 32,085,789-32,091,781): Nine sets of primers were designed to amplify an average of 800 bp spanning exons 2 to 9 and most introns.

Other genes (see Figure 2.1): Primers were designed to amplify one exon from each gene using 'PrimerQuest' [169].

Successful amplification was observed in 95% of PCRs. Failure was limited to a subset of strains and primer pairs and was most likely due to the presence of mismatches in the primer sequence. PCR reactions contained: 1.5 - 2 mM $MgCl_2$, 0.2 - 0.25 mM dNTPs, 0.2 - 1.8mM of each primer and 0.5 - 1 units of Taq polymerase (Promega) or Platinum Taq DNA Polymerase High Fidelity (Invitrogen) in a final volume of 10 - 50mL. Cycling conditions were 94°C, 4 min, 35 cycles at 94°, 55° and 68 - 72°C for 30 seconds each with a final extension at 68-72°C, 7 min. PCR products were purified using the High Pure PCR Product Purification kit (Roche) or fragments were excised from the gel, and purified using the Qiaquick gel extraction kit (Qiagen). Sequencing was performed at the UNC-CH Automated DNA Sequencing Facility on an ABI Prism 3730 (Applied Biosystems) or at the University of

Pennsylvania DNA Sequencing Facility using the Big Dye Terminator kit on an ABI 3730-48 capillary sequencer.

b) Sequence alignment and SNP identification and validation

All sequences were initially aligned using the Sequencher (Gene Codes) software. More refined alignment was done by eye to minimize, first, the number of events and second the number of variant positions. Aligned sequences were trimmed to retain only high-quality sequences. One fragment was eliminated from the analysis because of the presence of intra-strain polymorphisms consistent with the existence of a duplication. Analysis for sequence diversity was performed in the regions between the first and last nucleotide of high-quality sequence characterized in all 22 strains. Validation of the genetic variants was performed using several approaches: 1) Most insertion/deletion polymorphisms were confirmed in standard denaturing polyacrylamide gels; 2) SNPs detected among C57BL/6J, A/J, DBA/2J, and 129X1/SvJ were validated using the Celera Discovery System database [170]; 3) Some variants generated a restriction endonuclease cleavage site, and 27 of these variants were confirmed by restriction digestion and electrophoresis; and 4) Two SNPs were confirmed using Lightyper Simple Probe technology (Roche) designed to discriminate between the two alleles.

c) SDP analysis

Alleles at diallelic variants present in the 20 *M. musculus* strains were represented as a series of 0s and 1s in the following order: CAST, CASA, CALB JF1, MOLC, CZECH1, PWK, SKIVE, PERA, PERC, ZALENDE, TIRANO, LEWES, RBA, DDK, BALB/c, C57BL/6, DBA/2, A, and 129X1/Sv. The allele present in the first strain, CAST, is always a 0. Strains with the same allele as CAST were considered 0s and strains with different allele were considered 1s. Missing data were treated to minimize SDP number and to maximize SDP frequency within fragments.

d) Phylogenetic analysis

To date the insertion/deletion and substitution variants found among *M. musculus* inbred strains, we used the phylogenetic tree reported in Guenet and Bonhomme [109]. Dilallelic variants were classified as arising before the divergence of the specific lineages and before or after the divergence of the *M. musculus* subspecies (a, b and c respectively in Figure 2.5) using the following steps: 1) Variants at which SPRET and PANCEVO have different alleles were assigned to class "a". 2) Variants at which SPRET and PANCEVO share the same allele, this allele was considered ancestral. 3) Within this class, variants were assigned to class "b" if the new allele was found in inbred strains belonging to more than one subspecies. 4) Variants were assigned to class "c" if the new allele was found in inbred strains belonging to only one subspecies. To avoid biases due to the use of hybrid inbred strains derived from more than one subspecies, the following caveats were applied from step two onward. i) Classical inbred strains were omitted. ii) For wild-derived hybrid strains, alleles present at each variant were grouped with the appropriate parental subspecies in order to minimize the number of subspecies in which the new allele is found. In other words, alleles at each variant in JF1 and MOLC were classified as *M. m. castaneus* or *M. m. musculus*; alleles in SKIVE, were classified as *M. m. musculus* or *M. m. domesticus* and alleles in CALB, were classified as *M. m. castaneus* or *M. m. domesticus*. Lastly, if the allele present in either SPRET or PANCEVO was unknown, only variants in which both alleles were found in at least two subspecies each were considered to predate the divergence of the subspecies.

Discrimination between insertions and deletions was based on the predicted ancestral allele in the ancestor of the three *Mus* species analyzed here.

III. Mapping the paternal gene

a) Sequence variants

All base pair positions listed in Chapter 3 are based on the Ensembl v31 of the NCBI m33 mouse genome assembly [73]. Sequence variants in fragments 14-20 (positions 82,550,324-82,680,304) and 22-36 (positions 82,781,486-82,993,807), in the *Cct6b-Ap2b1* region, were included in the association analysis. In addition, we sequenced an 825 bp fragment in the intergenic region between *Slfn10* and *Slfn2* (positions 82651885 to 82652710; Forward 1: 5'-TCCAGATGTATATAATTGAGGTG-3', Forward 2: 5'-GACCCAGAATAAGGCCTAAC-3', Reverse 1: 5'-CAAGCTGAAGATTTGAAATGAC-3', Reverse 2: 5'-GTGCAAGTTAGAAACTACAAGC-3'; two fragments in the *Slfn1* gene: a 677 bp fragment in exon 1 (positions 82,729,880 to 82,730,557; Forward 1: 5'-GCTTCAACAGGTGCTCATGC-3', Forward 2: 5'-GGGGATTTGAGTGACGCTG-3', Reverse 1: 5'-CAGAGTTCCTAGAGAAGCACC-3', Reverse 2: 5'-CATTCTTGGTAGGCACTGG-3') and a 1193 bp in exon 2 (positions 82,734,269 to 82,735,462; Forward 1: 5'CAAGCACATTTTGCTGTTAGC-3', Forward 2: 5'-TGGAAAATGAACATCACCG-3', Forward 3: 5'-CCCAACACTAATGTCTCTGTC-3', Reverse 1: 5'-CAACATCCCCAGCTAAACG-3', Reverse 2: 5'-AAGACATGAGGAGCTTGATCC-3') (in each case, multiple sets of primers were used to ensure complete coverage). Finally, we have also sequenced a 341 bp fragment, from 83,005,403 to 83,004,744, distal to the *Ap2b1* gene (Forward: 5'-AAAGGGCGACTGACCCTCATCAAA-3', and Reverse: 5'-ATGGGTGGAGCACCAAACTACGTA-3').

The genotypes in C3H, WSB, KK, CBA and PWK at SNPs 82,843,176 and 82,843,476 were determined by PCR sequencing using the *Cct6b-Ap2b1*-26-F and R primers, 5'-CCTGCAATTTTCCTCTACAGC-3' and 5'-CCATGGAGATCTCTGTCCTG-3', respectively. The subspecific origin of the haplotypes among classical and hybrid strains in the *D11Spn173-D11Spn129* region (Table 3.2) was determined using diagnostic alleles at 26 variants distributed across the following fragments: *Cct6b-Ap2b1-25-27* (positions

106

82,833,035-82,857,930) and *Cct6b-Ap2b1-36* (positions 82,993,353-82,993,807) (Figure

6.1). Sequences have been submitted to EMBL and access numbers will be provided.

Strain	Type of strain	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	*M. m. castaneus*	*M. m. musculus*	*M. m. domesticus*	Subspecific origin
1. CAST	*M. m. castaneus*	A	C	C	C	A	**T**	A	C	A	G	T	G	**-**	T	A	A	**G**	C	G	T	G	T	G	**T**	G	C				
2. CASA	*M. m. castaneus*	A	C	C	C	A	**T**	A	C	A	G	T	G	**-**	T	A	A	**G**	C	G	T	G	T	G	**T**	G	C				
3. CALB	*M. m. castaneus*	A	C	C	C	A	**T**	A	C	A	G	T	G	**-**	T	A	A	**G**	C	G	T	G	T	G	**T**	G	C				
4. CZECH1	*M. m. musculus*	A	C	C	C	**T**	C	**G**	C	A	**C**	**G**	G	T	**G**	A	A	A	C	G	T	G	T	G	C	**A**	C				
5. PWK	*M. m. musculus*	A	C	C	C	**T**	C	**G**	C	A	**C**	**G**	G	T	**G**	A	A	A	C	G	T	G	T	G	C	**A**	C				
6. PERA	*M. m. domesticus*	**T**	**T**	C	**T**	A	C	A	C	**G**	**G**	T	G	T	T	**G**	**G**	A	C	**A**	**G**	**A**	**C**	G	C	G	**T**				
7. RBA	*M. m. domesticus*	**T**	**T**	C	**T**	A	C	A	C	**G**	**G**	T	G	T	T	**G**	**G**	A	C	**A**	**G**	**A**	**C**	A	C	G	**T**				
8. LEWES	*M. m. domesticus*	A	C	**T**	C	A	C	A	**-**	A	G	T	**T**	T	T	A	A	A	**T**	**A**	G	G	**C**	G	C	G	**T**				
9. PERC	*M. m. domesticus*	A	C	C	C	A	C	A	**-**	A	G	T	**T**	T	T	A	A	A	C	G	T	G	**C**	G	C	G	**T**				
10. ZALENDE	*M. m. domesticus*	A	C	C	C	A	C	A	**-**	A	G	T	**T**	T	T	A	A	A	C	G	T	G	**C**	G	C	G	**T**				
11. TIRANO	*M. m. domesticus*	A	C	C	C	A	C	A	**-**	A	G	T	**T**	T	T	A	A	A	C	G	T	G	**C**	G	C	G	**T**				
12. JF1	Hybrid	A	C	C	C	A	**T**	A	C	A	G	T	G	**-**	T	A	A	**G**	C	G	T	G	T	G	**T**	G	C	4	0	0	*castaneus*
13. MOLC	Hybrid	A	C	C	C	A	**T**	**G**	C	A	**C**	**G**	G	T	**G**	A	A	A	C	G	T	G	T	G	C	**A**	C	1	5	0	*musculus*
14. SKIVE	Hybrid	A	C	C	**T**	**T**	C	**G**	C	A	**C**	**G**	G	T	**G**	A	A	A	C	G	T	G	T	G	C	**A**	C	0	6	1	*musculus*
15. DDK	Classical	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	T	G	**C**	G	C	G	**T**	0	0	2	*domesticus*
16. BALB/c	Classical	**T**	**T**	C	**T**	A	C	A	C	**G**	G	T	G	T	T	**G**	**G**	A	C	**A**	**G**	**A**	**C**	A	C	G	**T**	0	0	12	*domesticus*
17. C57BL/6	Classical	**T**	**T**	C	**T**	A	C	A	C	**G**	G	T	G	T	T	**G**	**G**	A	C	**A**	**G**	**A**	**C**	A	C	G	**T**	0	0	12	*domesticus*
18. DBA/2	Classical	**T**	**T**	C	**T**	A	C	A	C	**G**	G	T	G	T	T	**G**	**G**	A	C	**A**	**G**	**A**	**C**	A	C	G	**T**	0	0	12	*domesticus*
19. A	Classical	**T**	**T**	C	**T**	A	C	A	C	**G**	G	T	G	T	T	**G**	**G**	A	C	**A**	**G**	**A**	**C**	G	C	G	**T**	0	0	11	*domesticus*
20. 129X1	Classical	**T**	**T**	C	**T**	A	C	A	C	**G**	G	T	G	T	T	**G**	**G**	A	C	**A**	**G**	**A**	**C**	G	C	G	**T**	0	0	11	*domesticus*
21. C3H	Classical	**T**	**T**	C	**T**	A	C	A	C	**G**	G	T	G	T	T	**G**	**G**	A	C	**A**	**G**	**A**	**C**	G	C	G	**T**	0	0	11	*domesticus*
22. CBA	Classical	**T**	**T**	C	**T**	A	C	A	*N*	*N*	G	T	G	T	T	**G**	**G**	A	C	**A**	**G**	*N*	**C**	G	C	G	**T**	0	0	9	*domesticus*
23. KK	Classical	A	C	**T**	C	A	C	A	**-**	A	G	T	**T**	T	T	A	A	A	**T**	**A**	G	G	**C**	G	C	G	**T**	0	0	8	*domesticus*

FIGURE 6.1. Phylogenetic origin of the candidate region in classical and hybrid strains.
The table provides the variants for which one allele is present only in strains from a single subspecies (*i.e.*, diagnostic alleles). Twenty-six variants containing diagnostic alleles were identified based on the distribution of alleles in wild-derived strains (strains 1-11) that can be unambiguously assigned to one of three *M. musculus* subspecies (*M. m. castaneus*, strains 1-3, variants 6, 13, 17 and 24; *M. m. musculus*, strains 4-5, variants 5, 7, 10-11, 14 and 25, and *M. m. domesticus*, strains 6-11, variants 1-4, 8-9, 12, 15, 16, 18-23 and 26). The diagnostic alleles at these variants were used to assign the haplotypes of three hybrid strains (strains 12-14) and nine classical strains (strains 15-23) to a *M. musculus* subspecies. The table also provides the number of diagnostic *castaneus*, *musculus* and *domesticus* alleles observed in each strain. Diagnostic alleles are shown in red boldface and underlined and in pink background. Italics denote unknown genotypes (N) or the presence of uninformative alleles. Deletions are denoted by (-). The variants are located in the following fragments: *Cct6b-Ap2b1-*25, variants 1-18; *Cct6b-Ap2b1-*26, variants 19-20; *Cct6b-Ap2b1-*27, variant 21 and *Cct6b-Ap2b1-*36, variants 22-26.

b) Selection of males for recombinant progeny testing

Over 5000 meioses from B6 x DDK backcrosses and intercrosses were screened for the presence of recombination between *D11Mit33* and *D11Mit35*. Animals carrying recombinant chromosomes were backcrossed to the parental strains or intercrossed to other B6 - DDK mixed background mice to generate 40 experimental males. Recombinants were genotyped at the following 15 informative markers: *D11Mit33*, *D11Mit35*, *D11Mit37*, *D11Mit66*, *D11Mit97*, *D11Mit120*, *D11Mit283*, *D11Mi354*, *D11Pas18*, *D11Spn1*, *D11Spn2*, *D11Spn4*, *Cct6b-Ap2b1-29*, *Scya1* and *Scya2* [56, 71, 91, 171]. In addition, recombinants were genotyped at 13 new microsatellite markers generated in our laboratories: *D11Spn10*, *D11Spn31*, *D11Spn36*, *D11Spn39*, *D11Spn72*, *D11Spn78*, *D11Spn104*, *D11Spn128*, *D11Spn129*, *D11Spn173*, *D11Spn178*, *Scya6* and *Scya7* (Table 6.1). Accession numbers at EMBL Nucleotide Sequence Database are AM075608 – AM075620. Genotypes were determined by PCR amplification and electrophoresis as described previously [55, 56, 69, 75, 90]. Only markers flanking the recombination events in the candidate region were used in the mapping analyses to maximize the information with the fewest number of markers (Table 3.1a-b).

Inbred males: Wild-derived strains were chosen to include a broad spectrum of *M. musculus* subspecies and a related species, *M. spicilegus*. We included an excess of *M. m. domesticus* strains because this is the only subspecies for which wild-derived strains have not been characterized previously for compatibility with the maternal DDK factor [74]. The classical strains included in this study are commonly used in mouse genetics studies.

TABLE 6.1. Sequence and position of oligonucleotides used to genotype novel microsatellites in the *Om* region.

Marker	Ensembl position (NCBI Build 33)		Primer F	Primer R	B6	DDK
	bp start	bp end				
Scya7	81660424	81660632	GAAATGTGCCTGAACAGAAACC	TTCAAATCACACCAAAGTACATGG	209	<209
D11Spn36	82514021	82514173	GATGGACCTTCTCAGAGTCCC	CAAGGGTCAGCTGTGAAGGCC	153	>153
D11Spn39	82514117	82514388	CCGCAGATGGAGAAGATGACTC	GTTATTTATTTGGACACACGC	272	>272
D11Spn173	82560290	82560436	CATATGCACACATACACAATACG	CCATTCATTTTAAAGTAACTGGG	169	167
D11Spn178	82585484	82585662	CTCCTTCAAAGGACAAAGTGC	GTTCATGTTCAACTCCATTCTC	179	<179
D11Spn31	82605689	82605810	CCAGATTGTTCAGTGTTGAATG	TCTATCCACTAGGTGAAGTGG	142	140
D11Spn78	82651101	82651348	TCTGAAGAGAGCGACAGTGTG	CTGACATCTATGAGCCAAAGCG	248	<248
D11Spn72	82828247	82828418	TCTGGAGCTGTTGAAAAGGATG	CAATCGGAGCCTTGCTCCTACC	174	<174
D11Spn128	82970687	82970864	CTCACAGACAAGAGGGAGATTC	CTGGAGGGACAGATTCAATGAG	178	<178
D11Spn129	83025035	83025266	GATTATCCTGAGTCCCTTTTGAG	CTCTGAGATGGAGTTGGTTTCTC	232	>232
D11Spn104	83047325	83047464	TGGCAGCAAGGACAAGCACC	CATACGTGAGCATACATGCAA	140	>140
Scya6	83192067	83192187	GATTTCCTCTTGCACGCAGCG	TATCTGGCATAAGAGACAGG	121	>121
D11Spn10	83245571	83245742	AACCCAACCACCAAGACTTAAGAG	TGTATATTTTAAGTAAATGCCTGG	172	<172

The table provides the allele sizes for the two strains used in the progeny testing analysis.

c) Reproductive performance

In all crosses described in this study, the dam is always listed first and the sire second unless otherwise indicated. Males were mated to 9-45 weeks old identical (B6 x DDK)F$_1$ females. The three crosses used as controls in this study, (B6 x DDK)F$_1$ x C57BL/6, (B6 x DDK)F$_1$ x DDK and (B6 x DDK)F$_1$ x (B6 x DDK)F$_1$ have been described previously [56, 66, 69, 75, 90, 172]. Some of the crosses involving inbred males have been previously described in our efforts to characterize a meiotic drive phenotype [91, 92]. Briefly, each male was mated to identical (B6 x DDK)F$_1$ females and monitored daily for newborn pups. The

litter size was determined at birth to avoid biases due to postnatal lethality and cannibalization. The distribution of litter sizes produced by each type of male was used to determine the phenotype of the male.

Correction of reproductive performance: Lower reproductive performance has been observed consistently in inter-specific and inter-subspecific crosses, due to embryonic lethality that is not associated with the DDK syndrome (Alibert et al. 1997; Britton-Davidian et al. 2005). Therefore, it is necessary to correct for this effect based on the estimated phylogenetic divergence between the dam and the sire. Because the females used in this study are hybrids of two classical inbred strains, whose genomes are derived mostly from *M. m. domesticus* no correction is necessary in crosses involving males from classical strains, *M. m. domesticus* wild-derived strains or JF1 (see above). A second group includes a single strain, SKIVE, a hybrid of *M. m. musculus* and M. m. domesticus and the correction factor used was 1.09 (Table 6.2). Lastly, a correction factor of 1.29 was used in crosses involving males from other subspecies (MOLC and CAST) and *M. spicilegus* (PANCEVO). These correction factors were determined experimentally in crosses that cannot have DDK syndrome related lethality and are the ratio between the average litter size in intra-subspecific crosses and the average litter size in inter-specific and inter-subspecific crosses (Table 6.2). The corrected values are used throughout our analyses.

d) Dot plot matrix

The sequence flanked by *D11Spn173* and *D11Spn129* (positions 82,560,436 to 83,025,035 on chromosome 11) was retrieved from the Ensembl gene build for the NCBI m33 mouse assembly [73]. The homologous region in the rat genome (positions 71,159,554 to 71,507,449 on chromosome 10, version 3.4 "November 2004 Update" of the rat genome assembly) was identified by BLAST search using the first and last 1 kb of the mouse

TABLE 6.2. Crosses used to determine the correction factors for reproductive performance in interspecific (Is) and intersubspecific (Iss) crosses

Type of Cross	Dam	Sire	Average litter size	SD	No of litters	References
Is	(B6 x CASP)F$_1$	CASP	7.30	1.20	9	[61]
	(B6 x MOM)F$_1$	B6	6.30	1.66	11	[61]
	(B6 x MOM)F$_1$	MOM	7.10	1.73	12	[61]
	(CAST x B6)F$_1$	B6	8.27	1.57	66	‡
	B6	(B6 x CASP)F$_1$	7.90	1.44	13	[61]
	B6	(B6 x MOM)F$_1$	7.30	2.71	15	[61]
	C3H	(C3H x MOM)F$_1$	8.00	1.32	7	[61]
	Subtotal		7.45 (1.29)			
Hybrid	(B6 x SKIVE)F$_1$	B6	8.74	1.79	39	‡
	Subtotal		8.74 (1.09)			
Iss	(PERA x B6)F$_1$	B6	9.90	0.94	11	‡
	(B6 x PERA)F$_1$	B6	10.04	1.95	104	‡
	(B6 x PERC)F$_1$	B6	8.9	2.33	50	‡
	Subtotal		9.61 (1.00)			

The correction factors used for interspecific and hybrid crosses are the ratio between the average litter size in intrasubspecific crosses and the average litter size in intersubspecific and hybrid crosses, respectively, and are shown in parenthesis.
‡This study.

interval. Dot plot analyses (Maizel and Lenk 1981; Pustell and Kafatos 1982; Quigley et al. 1984) were used to visualize the presence of duplications and inversions within *M. musculus* and to align homologous regions of the mouse and rat. These analyses were performed using a dedicated website at the University of Colorado [173]. We used a 99 bp sliding window and 10 bp mismatch limit in the production of the matrix plots. The position and identity of genes in the mouse and rat regions was retrieved from Ensembl [73, 174], UCSC [175] and NCBI [176]. In addition, mouse genes were used to identify additional putative homologous genes in the rat using BLAST search.

e) Phylogenetic analyses

Sequences from the *Slfn1* gene and fragments *Cct6b-Ap2b1-*22 - 27 (positions

82,781,486-82,824,604 and 82,857,332-82,857,930) were aligned. Fragment Cct6b-Ap2b1-

25 was excluded from these analyses because the DDK strain has a deletion encompassing

this fragment. In addition to the 17 inbred strains phenotyped, we also included PWK and

SPRET to ensure complete representation of different lineages. All analyses were

performed using the PHYLIP phylogeny inference software package, version 3.6 [177]. For

each set of sequences we generated 100 bootstrapped datasets using the SEQBOOT

program. We then determined the phylogeny using a distance matrix method (NEIGHBOR),

a maximum likelihood method (DNAML) and a maximum parsimony method (DNAPARS).

The CONSENSE program was used to construct majority rule consensus trees.

f) Statistical analyses and mapping

Litter size was analyzed using mixed and nested-mixed models within the computer

program SAS version 9.1 (SAS Institute, Cary, NC). Three sets of statistical analyses were

performed. In all analyses, to account for the correlation between litter sizes from the same

male, 'male' was treated as a random effect in our statistical models. In the first set of

analyses, we tested for associations between genotype and litter size in 40 males with

different B6 - DDK backgrounds using a mixed model approach. In our second set of

statistical models, we calculated the average litter size and their associated 99% confidence

intervals for each of the 17 strains phenotyped. After accounting for the father, it was

determined that the correlation in litter size due to the dam was not statistically significant.

Finally, in our third set of statistical models, we analyzed 167 markers in 17 inbred strains to

determine whether genotype was associated with litter size using nested mixed models.

Genotype was treated as a fixed effect while strain was treated as a random effect, to

account for correlation of litter sizes within strains, and father's identity was treated as a

nested random effect within strain. After accounting for genotype, the residuals were distributed in a symmetric, unimodal fashion, consistent with distributional assumptions necessary for the validity of the mixed model. The distribution of litter size was ordinal, with values ranging from 1 to 16. Given the discrete ordinal nature of the data, assumptions regarding the normality of the residuals in our statistical models were of some concern. However the large sample size and the resulting asymptotics should make our p-values fairly robust. To validate our findings we performed permutation tests, based on 10,000 to 100,000 random replicate data sets, to assess the empirical statistical significance of our findings. For this application, permutation tests are very conservative because in order to maintain the correlation structure of the data and hence 'exchangeability', blocks consisting of all the litters for a given male (for the first set of analyses) or a given strain (for the third set of analyses) were permuted together rather than permuting individual litters. This constraint led to a relatively small number of possible permutations of the data and hence a conservative lower bound on the p-values. In some instances it was feasible to calculate the empirical probability of our findings under the null hypothesis by determining the exact number of possible permutations of the genotype data that would result in a as extreme or more extreme F statistics than the ones observed in our mixed models. In addition to protecting our conclusions from inaccurate p-values due to erroneous distributional assumptions, the permutation test procedure offers some protection from over interpreting results from a relatively small sample of males or strains, as the statistical tests are conditional on the observed phenotype and genotype data.

IV. Rescue modifiers of the DDK syndrome

a) Mating schemes

For all crosses described in this study the dam is listed first and the sire second, unless otherwise indicated. F_1 hybrid females were generated by crossing DDK males to B6, 129X1, DBA/2, BALB/c, JF1, CAST, WSB, PERC, PERA, RBA or SKIVE females. PERC males were also crossed to DDK/Pas females to generate reciprocal F_1 females. The reproductive performance of (B6 x DDK)F_1 x B6 and (B6 x DDK)F_1 x DDK crosses were described previously [55, 69, 75]. F_2 males were generated in a (PERC x B6)F_1 intercross. Male offspring were selected based on their genotypes at *Om* and *Rmod1*. All crosses involved females between the ages of 2-10 months and males between the ages of 2-12 months. Cages were checked daily for the presence of newborn pups and litter size was recorded to avoid biases due to postnatal lethality.

b) Reproductive performance and statistical analysis

The reproductive performance of each cross was estimated using the litter size as described previously [55, 69, 75]. Average litter sizes of zero were accounted for by dividing the sum of the females' average litter sizes by the number of females. Unless otherwise noted, experimental crosses involved a dam and sire with incompatible alleles at the gene encoding the maternal factor and the paternal gene, respectively, and control crosses had a compatible combination of alleles at *Om*. The level of rescue in the F_1 hybrid screen was determined by comparing the distribution of litter sizes of experimental crosses to that of control crosses. Under the null hypothesis that these two values are equal, significance was calculated using the Wilcoxon signed-rank test (JMP 6 release 6.0, S.A.S. Institute, Cary, N.C.).

In tests to determine whether transmission through the sire of PERA or PERC alleles at *Rmod1* or DDK alleles at *Rmod2* leads to rescue; and when determining if the presence of two B6 alleles at rescue modifier loci is associated with a decrease in lethality, statistical significance was calculated using oneway analysis of variance (JMP 6 release 6.0, S.A.S. Institute, Cary, N.C.) under the null hypothesis that experimental and control crosses have equal reproductive performance.

In tests to determine whether $Om^{B6/DDK}$ females that are homozygous for B6 alleles at *Rmod1* or *Rmod2* have significantly different reproductive performance when compared to $Om^{B6/DDK}$ females that are heterozygous B6/DDK at *Rmod1* or *Rmod2*, statistical significance was calculated using oneway analysis of variance (JMP 6 release 6.0, S.A.S. Institute, Cary, N.C.) under the null hypothesis that both types of females have equal reproductive performance. Where noted, *p*-value significance was corrected for multiple testing using the Bonferroni correction (alpha (0.05) divided by the number of tests done (3)).

c) Genotyping

All base pair positions listed in Chapter 4 are based on the Ensembl v42 of the NCBI m36 mouse genome assembly [73]. Genotyping, by PCR amplification and gel electrophoresis, was performed on DNA obtained from The Jackson Laboratory (Bar Harbor, Me, USA); or on DNA extracted from tail biopsies as described previously [92]. Oligonucleotide primers for microsatellite markers [171] were either purchased from Integrated DNA Technologies [169] or Invitrogen Research Genetics [178]. Alleles at *Om* were inferred from genotypes at markers within (*D11Spn31* or *D11Spn78*) or flanking (*D11Mit35* and *D11Mit33*) the candidate interval for the maternal gene (see Chapter 3). Marker *D13Mit135* (position 21,933,671) was used to infer genotypes at *Rmod1*.

d) Genome scan

A whole genome scan was performed on 36 heterozygous Om^{DDK}/Om^{B6} G_2 females having B6-DDK-PERA mixed backgrounds and 39 heterozygous Om^{DDK}/Om^{B6} G_2 females having B6-DDK-PERC mixed backgrounds. DNA from these females was extracted from tail biopsies using the GenElute^{TM} Mammalian Genomic DNA Miniprep Kit (SIGMA, [179]. Genotypes were obtained by one of three methods: 1) In an initial whole genome scan we genotyped 394 SNPs using Sequenom MassARRAY technology followed by refinement on chromosome 13 by genotyping 34 additional SNPs using Sequenom Iplex technology [180] [181]; 2) Additional genome scan in the PERC crosses with a panel of 768 SNPs genotyped by Illumina (San Diego, CA) using the BeadArray technology [182]; and 3) Microsatellite genotyping using PCR and acrylamide gel electrophoresis were used to fill in large gaps between markers. Overall, marker coverage spanned 19 chromosomes (markers on chromosome 11 were excluded from this analysis because all G_2 females were selected to be Om^{DDK}/Om^{B6}). Additional markers were excluded on the basis of: 1) our inability to score genotypes in more than 50% of the samples, 2) failure of the assay to recognize both of the two possible genotypes, 3) the presence of impossible genotypes, and 4) genotypes that created impossible recombination events between closely linked markers. On average, markers were spaced every 3 cM (range 0-25 cM) in PERC crosses, or every 10 cM (range 0-35 cM) in PERA crosses.

e) Linkage Analysis

Genetic distances of informative SNPs included in this analysis were inferred using closely linked microsatellite markers for which the genetic distance from the centromere was determined previously [183]. We used the average litter size as the phenotype for QTL mapping (similar results were obtained when the median litter size was used). Linkage

116

analyses were performed using a non-parametric model (an extension of the Kruskal-Wallis test) within the R/qtl software program [184]. This approach yielded more conservative linkage values and was better suited to our data compared to the parametric model within R/qtl (expectation-maximization), because the average litter sizes were not normally distributed. Using 1000 permutations within R/qtl, we determined that the genome-wide significance thresholds while p=0.05 are: 2.48 in PERC crosses, and 2.31 in PERA crosses; and while p=0.01 are: 3.05 in PERC crosses and 3.31 in PERA crosses. The percent variance explained by each QTL was estimated using the fitqtl function of R/qtl. Markers *D13Mit135* and *rs13479321* were used to infer the effects of *Rmod1* and *Rmod2*.

f) Phylogenetic analyses

SNPs from the GNF panel within the candidate interval for *Rmod1* (positions 19,772,465 – 26,254,479) were aligned in Phylip format [185]. All analyses were performed using the PHYLIP phylogeny inference software package, version 3.6 [177]. To test for SNP selection biases in the GNF dataset, we compared phylogenetic analysis (maximum parsimony, DNAPARS) of SNP data in 1 Mb intervals to the phylogenetic analysis based on the Perlegen datasets [153]. This was done using strains that were present in both datasets. This comparison showed that the resulting phylogenetic relationships between the strains based on the GNF dataset is relatively consistent over the entire 6.5 Mb interval and is mostly the same as predicted by the Perlegen data.

The consensus tree depicted in Figure 5.3 was generated using 100 bootstrapped datasets in the SEQBOOT program. We then determined the phylogeny using a distance matrix method (NEIGHBOR), a maximum likelihood method (DNAML), and a maximum parsimony method (DNAPARS). The CONSENSE program was used to construct majority

rule consensus trees. The phylogram was constructed using a maximum parsimony method (DNAPARS).

g) Gene expression profiles at Rmod1

Expression of genes within the *Rmod1* candidate interval was determined by comparing genomic sequence to the mouse EST libraries available through NCBI [186]. Sequences from the 6.5Mb candidate interval for *Rmod1* were blasted in consecutive 250 kb intervals using the keyword oocyte. In addition, the gene expression profile of the 134 known genes within the interval [73] were determined based on Unigene [186].

BIBLIOGRAPHY

1.   Edwards RG. Genetics, epigenetics and gene silencing in differentiating mammalian embryos. Reprod Biomed Online 2006; 13: 732-753.

2.   Lu CC, Brennan J, Robertson EJ. From fertilization to gastrulation: axis formation in the mouse embryo. Curr Opin Genet Dev 2001; 11: 384-392.

3.   Reik W, Santos F, Mitsuya K, Morgan H, Dean W. Epigenetic asymmetry in the mammalian zygote and early embryo: relationship to lineage commitment? Philos Trans R Soc Lond B Biol Sci 2003; 358: 1403-1409; discussion 1409.

4.   Latham KE, Sapienza C. Developmental potential as a criterion for understanding and defining embryos. Conn Law Rev 2004; 36: 1171-1176.

5.   Munne S. Preimplantation genetic diagnosis of numerical and structural chromosome abnormalities. Reprod Biomed Online 2002; 4: 183-196.

6.   Ola B, Li TC. Implantation failure following in-vitro fertilization. Curr Opin Obstet Gynecol 2006; 18: 440-445.

7.   O'Neill C. A consideration of the factors which influence and control the viability and developmental potential of the preimplantation embryo. Baillieres Clin Obstet Gynaecol 1991; 5: 159-178.

8.   Lucifero D, Chaillet JR, Trasler JM. Potential significance of genomic imprinting defects for reproduction and assisted reproductive technology. Hum Reprod Update 2004; 10: 3-18.

9.   Kwong WY, Wild AE, Roberts P, Willis AC, Fleming TP. Maternal undernutrition during the preimplantation period of rat development causes blastocyst abnormalities and programming of postnatal hypertension. Development 2000; 127: 4195-4202.

10.   Buiting K, Dittrich B, Gross S, Lich C, Farber C, Buchholz T, Smith E, Reis A, Burger J, Nothen MM, Barth-Witte U, Janssen B, Abeliovich D, Lerer I, van den Ouweland AM, Halley DJ, Schrander-Stumpel C, Smeets H, Meinecke P, Malcolm S, Gardner A, Lalande M, Nicholls RD, Friend K, Schulze A, Matthijs G, Kokkonen H, Hilbert P, Van Maldergem L, Glover G, Carbonell P, Willems P, Gillessen-Kaesbach G, Horsthemke B. Sporadic imprinting defects in Prader-Willi syndrome and Angelman

syndrome: implications for imprint-switch models, genetic counseling, and prenatal diagnosis. Am J Hum Genet 1998; 63: 170-180.

11.   Thompson JG, Kind KL, Roberts CT, Robertson SA, Robinson JS. Epigenetic risks related to assisted reproductive technologies: short- and long-term consequences for the health of children conceived through assisted reproduction technology: more reason for caution? Hum Reprod 2002; 17: 2783-2786.

12.   Reik W, Romer I, Barton SC, Surani MA, Howlett SK, Klose J. Adult phenotype in the mouse can be affected by epigenetic events in the early embryo. Development 1993; 119: 933-942.

13.   Song JL, Wessel GM. How to make an egg: transcriptional regulation in oocytes. Differentiation 2005; 73: 1-17.

14.   Wassarman PM, Kinloch RA. Gene expression during oogenesis in mice. Mutat Res 1992; 296: 3-15.

15.   McGrath SA, Esquela AF, Lee SJ. Oocyte-specific expression of growth/differentiation factor-9. Mol Endocrinol 1995; 9: 131-136.

16.   Kurasawa S, Schultz RM, Kopf GS. Egg-induced modifications of the zona pellucida of mouse eggs: effects of microinjected inositol 1,4,5-trisphosphate. Dev Biol 1989; 133: 295-304.

17.   Xu Z, Kopf GS, Schultz RM. Involvement of inositol 1,4,5-trisphosphate-mediated Ca2+ release in early and late events of mouse egg activation. Development 1994; 120: 1851-1859.

18.   Swann K, Ozil JP. Dynamics of the calcium signal that triggers mammalian egg activation. Int Rev Cytol 1994; 152: 183-222.

19.   Swann K, Larman MG, Saunders CM, Lai FA. The cytosolic sperm factor that triggers Ca2+ oscillations and egg activation in mammals is a novel phospholipase C: PLCzeta. Reproduction 2004; 127: 431-439.

20.   Runft LL, Jaffe LA, Mehlmann LM. Egg activation at fertilization: where it all begins. Dev Biol 2002; 245: 237-254.

21.   Bleil JD, Wassarman PM. Mammalian sperm-egg interaction: identification of a glycoprotein in mouse egg zonae pellucidae possessing receptor activity for sperm. Cell 1980; 20: 873-882.

22.   Tsaadon A, Eliyahu E, Shtraizent N, Shalgi R. When a sperm meets an egg: block to polyspermy. Mol Cell Endocrinol 2006; 252: 107-114.

23.   Scott GF. Developmental Biology. Sunderland: Sinauer Associates, Inc.; 2000.

24.   Donahue RP. Cytogenetic analysis of the first cleavage division in mouse embryos (fertilization-pronuclei-T163H translocation). Proc Natl Acad Sci U S A 1972; 69: 74-77.

25.   Donahue RP. Fertilization of the mouse oocyte: sequence and timing of nuclear progression to the two-cell stage. J Exp Zool 1972; 180: 305-318.

26.   Gulyas BJ. A reexamination of cleavage patterns in eutherian mammalian eggs: rotation of blastomere pairs during second cleavage in the rabbit. J Exp Zool 1975; 193: 235-248.

27.   Gardner RL. Experimental analysis of second cleavage in the mouse. Hum Reprod 2002; 17: 3178-3189.

28.   Piko L, Clegg KB. Quantitative changes in total RNA, total poly(A), and ribosomes in early mouse embryos. Dev Biol 1982; 89: 362-378.

29.   Paynton BV, Rempel R, Bachvarova R. Changes in state of adenylation and time course of degradation of maternal mRNAs during oocyte maturation and early embryonic development in the mouse. Dev Biol 1988; 129: 304-314.

30.   Alizadeh Z, Kageyama S, Aoki F. Degradation of maternal mRNA in mouse embryos: selective degradation of specific mRNAs after fertilization. Mol Reprod Dev 2005; 72: 281-290.

31.   Bachvarova R, De Leon V. Polyadenylated RNA of mouse ova and loss of maternal RNA in early development. Dev Biol 1980; 74: 1-8.

32.   Flach G, Johnson MH, Braude PR, Taylor RA, Bolton VN. The transition from maternal to embryonic control in the 2-cell mouse embryo. Embo J 1982; 1: 681-686.

33.   Fleming TP, Javed Q, Hay M. Epithelial differentiation and intercellular junction formation in the mouse early embryo. Dev Suppl 1992: 105-112.

34.   Becker DL, Davies CS. Role of gap junctions in the development of the preimplantation mouse embryo. Microsc Res Tech 1995; 31: 364-374.

35.   Ducibella T, Albertini DF, Anderson E, Biggers JD. The preimplantation mammalian embryo: characterization of intercellular junctions and their appearance during development. Dev Biol 1975; 45: 231-250.

36.   O'Rahilly R. Early human development and the chief sources of information on staged human embryos. Eur J Obstet Gynecol Reprod Biol 1979; 9: 273-280.

37.   Theiler K. The House Mouse: development and Normal Stages from Fertilization to Four Weeks of Age. Verlag, NY: Springer; 1972.

38.   Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. Science 2001; 293: 1089-1093.

39.   Dean W, Santos F, Reik W. Epigenetic reprogramming in early mammalian development and following somatic nuclear transfer. Semin Cell Dev Biol 2003; 14: 93-100.

40.   Huynh KD, Lee JT. Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos. Nature 2003; 426: 857-862.

41.   Zuccotti M, Boiani M, Ponce R, Guizzardi S, Scandroglio R, Garagna S, Redi CA. Mouse Xist expression begins at zygotic genome activation and is timed by a zygotic clock. Mol Reprod Dev 2002; 61: 14-20.

42.   Sado T, Ferguson-Smith AC. Imprinted X inactivation and reprogramming in the preimplantation mouse embryo. Hum Mol Genet 2005; 14 Spec No 1: R59-64.

43.   Okamoto I, Otte AP, Allis CD, Reinberg D, Heard E. Epigenetic dynamics of imprinted X inactivation during early mouse development. Science 2004; 303: 644-649.

44.   Tamada H, Kikyo N. Nuclear reprogramming in mammalian somatic cell nuclear cloning. Cytogenet Genome Res 2004; 105: 285-291.

45. Dean W, Santos F, Stojkovic M, Zakhartchenko V, Walter J, Wolf E, Reik W. Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. Proc Natl Acad Sci U S A 2001; 98: 13734-13738.

46. Latham KE. Cloning: questions answered and unsolved. Differentiation 2004; 72: 11-22.

47. Falls JG, Pulford DJ, Wylie AA, Jirtle RL. Genomic imprinting: implications for human disease. Am J Pathol 1999; 154: 635-647.

48. Schofield PN, Joyce JA, Lam WK, Grandjean V, Ferguson-Smith A, Reik W, Maher ER. Genomic imprinting and cancer; new paradigms in the genetics of neoplasia. Toxicol Lett 2001; 120: 151-160.

49. Li G, Su Q, Liu GQ, Gong L, Zhang W, Zhu SJ, Zhang HL, Feng YM. Skewed X chromosome inactivation of blood cells is associated with early development of lung cancer in females. Oncol Rep 2006; 16: 859-864.

50. Babinet C, Richoux V, Guenet JL, Renard JP. The DDK inbred strain as a model for the study of interactions between parental genomes and egg cytoplasm in mouse preimplantation development. Dev Suppl 1990: 81-87.

51. Tomita T. One-side cross sterility between inbred strains of mice. Jpn J Genet 1960; 35: 291.

52. Wakasugi N, Tomita T, Kondo K. Differences of fertility in reciprocal crosses between inbred strains of mice. DDK, KK and NC. J Reprod Fertil 1967; 13: 41-50.

53. Wakasugi N. A genetically determined incompatibility system between spermatozoa and eggs leading to embryonic death in mice. J Reprod Fertil 1974; 41: 85-96.

54. Wakasugi N. Studies on fertility of DDK mice: reciprocal crosses between DDK and C57BL/6J strains and experimental transplantation of the ovary. J Reprod Fertil 1973; 33: 283-291.

55. Pardo-Manual de Villena F, Slamka C, Fonseca M, Naumova AK, Paquette J, Pannunzio P, Smith M, Verner A, Morgan K, Sapienza C. Transmission-ratio distortion through F1 females at chromosome 11 loci linked to Om in the mouse DDK syndrome. Genetics 1996; 142: 1299-1304.

56.   Pardo-Manuel De Villena F, de La Casa-Esperon E, Williams JW, Malette JM, Rosa M, Sapienza C. Heritability of the maternal meiotic drive system linked to Om and high-resolution mapping of the Responder locus in mouse. Genetics 2000; 155: 283-289.

57.   Renard JP, Babinet C. Identification of a paternal developmental effect on the cytoplasm of one-cell-stage mouse embryos. Proc Natl Acad Sci U S A 1986; 83: 6883-6886.

58.   Mann JR. DDK egg-foreign sperm incompatibility in mice is not between the pronuclei. J Reprod Fertil 1986; 76: 779-781.

59.   Cockett NE, Jackson SP, Shay TL, Farnir F, Berghmans S, Snowder GD, Nielsen DM, Georges M. Polar overdominance at the ovine callipyge locus. Science 1996; 273: 236-238.

60.   Georges M, Charlier C, Cockett N. The callipyge locus: evidence for the trans interaction of reciprocally imprinted genes. Trends Genet 2003; 19: 248-252.

61.   Wakasugi N, Morita M. Studies on the development of F1 embryos from inter-strain cross involving DDK mice. J Embryol Exp Morphol 1977; 38: 211-216.

62.   Buehr M, Lee S, McLaren A, Warner A. Reduced gap junctional communication is associated with the lethal condition characteristic of DDK mouse eggs fertilized by foreign sperm. Development 1987; 101: 449-459.

63.   Leclerc C, Becker D, Buehr M, Warner A. Low intracellular pH is involved in the early embryonic death of DDK mouse eggs fertilized by alien sperm. Dev Dyn 1994; 200: 257-267.

64.   Renard JP, Baldacci P, Richoux-Duranthon V, Pournin S, Babinet C. A maternal factor affecting mouse blastocyst formation. Development 1994; 120: 797-802.

65.   Gao S, Wu G, Han Z, de la Casa-Esperon E, Sapienza C, Latham KE. Recapitulation of the ovum mutant (Om) phenotype and loss of Om locus polarity in cloned mouse embryos. Biol Reprod 2005; 72: 487-491.

66.   Sapienza C, Paquette J, Pannunzio P, Albrechtson S, Morgan K. The polar-lethal Ovum mutant gene maps to the distal portion of mouse chromosome 11. Genetics 1992; 132: 241-246.

67.   Le Bras S, Cohen-Tannoudji M, Kress C, Vandormael-Pournin S, Babinet C, Baldacci P. BALB/c alleles at modifier loci increase the severity of the maternal effect of the "DDK syndrome". Genetics 2000; 154: 803-811.

68.   Baldacci PA, Richoux V, Renard JP, Guenet JL, Babinet C. The locus Om, responsible for the DDK syndrome, maps close to Sigje on mouse chromosome 11. Mamm Genome 1992; 2: 100-105.

69.   Pardo-Manuel de Villena F, Naumova AK, Verner AE, Jin WH, Sapienza C. Confirmation of maternal transmission ratio distortion at Om and direct evidence that the maternal and paternal "DDK syndrome" genes are linked. Mamm Genome 1997; 8: 642-646.

70.   Baldacci PA, Cohen-Tannoudji M, Kress C, Pournin S, Babinet C. A high-resolution map around the locus Om on mouse Chromosome 11. Mamm Genome 1996; 7: 114-116.

71.   Cohen-Tannoudji M, Vandormael-Pournin S, Le Bras S, Coumailleau F, Babinet C, Baldacci P. A 2-Mb YAC/BAC-based physical map of the ovum mutant (Om) locus region on mouse chromosome 11. Genomics 2000; 68: 273-282.

72.   Le Bras S, Cohen-Tannoudji M, Guyot V, Vandormael-Pournin S, Coumailleau F, Babinet C, Baldacci P. Transcript map of the Ovum mutant (Om) locus: isolation by exon trapping of new candidate genes for the DDK syndrome. Gene 2002; 296: 75-86.

73.   http://www.ensembl.org/Mus_musculus/index.html.

74.   Zhao WD, Ishikawa A, Yamagata T, Bolor H, Wakasugi N. Female mice of DDK strain are fully fertile in the intersubspecific crosses with Mus musculus molossinus and M. m. castaneus. Mamm Genome 2002; 13: 345-351.

75.   Pardo-Manuel de Villena F, de la Casa-Esperon E, Verner A, Morgan K, Sapienza C. The maternal DDK syndrome phenotype is determined by modifier genes that are not linked to Om. Mamm Genome 1999; 10: 492-497.

76.   Zhao WD, Ishikawa A, Wakasugi N. Fluctuation in fertility phenotypes of the heterozygous (Om/+) mice owing to background genes. Mamm Genome 2002; 13: 114-116.

77. Zhao WD, Chung HJ, Wakasugi N. Modification of survival rate of mouse embryos developing in heterozygous females for ovum mutant gene. Biol Reprod 2000; 62: 857-863.

78. Riviere I, Sunshine MJ, Littman DR. Regulation of IL-4 expression by activation of individual alleles. Immunity 1998; 9: 217-228.

79. Kelly BL, Locksley RM. Coordinate regulation of the IL-4, IL-13, and IL-5 cytokine cluster in Th2 clones revealed by allelic expression patterns. J Immunol 2000; 165: 2982-2986.

80. Rhoades KL, Singh N, Simon I, Glidden B, Cedar H, Chess A. Allele-specific expression patterns of interleukin-2 and Pax-5 revealed by a sensitive single-cell RT-PCR analysis. Curr Biol 2000; 10: 789-792.

81. Pastinen T, Sladek R, Gurd S, Sammak A, Ge B, Lepage P, Lavergne K, Villeneuve A, Gaudin T, Brandstrom H, Beck A, Verner A, Kingsley J, Harmsen E, Labuda D, Morgan K, Vohl MC, Naumova AK, Sinnett D, Hudson TJ. A survey of genetic and epigenetic variation affecting human gene expression. Physiol Genomics 2004; 16: 184-193.

82. Chess A. Monoallelic expression of protocadherin genes. Nat Genet 2005; 37: 120-121.

83. Balomenos D, Balderas RS, Mulvany KP, Kaye J, Kono DH, Theofilopoulos AN. Incomplete T cell receptor V beta allelic exclusion and dual V beta-expressing cells. J Immunol 1995; 155: 3308-3312.

84. Rassenti LZ, Kipps TJ. Lack of allelic exclusion in B cell chronic lymphocytic leukemia. J Exp Med 1997; 185: 1435-1445.

85. Wakui M, Kim J, Butfiloski EJ, Morel L, Sobel ES. Genetic dissection of lupus pathogenesis: Sle3/5 impacts IgH CDR3 sequences, somatic mutations, and receptor editing. J Immunol 2004; 173: 7368-7376.

86. Muller HJ. Isolating mechanisms, evolution and temperature. Biol. Symp. 1942; 6: 71-125.

87. Dobzhansky T. Studies on Hybrid Sterility. II. Localization of Sterility Factors in Drosophila Pseudoobscura Hybrids. Genetics 1936; 21: 113-135.

88. Sandler L, Novistshi E. Meiotic drive as an evolutionary force. American Naturalist 1957; 91: 105-110.

89. Pardo-Manual de Villena F, Sapienza C. Nonrandom segregation during the female meiosis: the unfairness of females. Mamm Genome 2001; 12: 331-339.

90. Pardo-Manuel de Villena F, de la Casa-Esperon E, Briscoe TL, Sapienza C. A genetic test to determine the origin of maternal transmission ratio distortion. Meiotic drive at the mouse Om locus. Genetics 2000; 154: 333-342.

91. Wu G, Hao L, Han Z, Gao S, Latham KE, de Villena FP, Sapienza C. Maternal transmission ratio distortion at the mouse Om locus results from meiotic drive at the second meiotic division. Genetics 2005; 170: 327-334.

92. Kim K, Thomas S, Howard I, Bell TA, Doherty HE, Ideraabdullah F, Detwiler DA, Pardo-Manuel de Villena F. Meiotic drive at the Om locus in wild-derived inbred mouse strains. Biol J Linn Soc 2005; 84: 487-492.

93. Beck JA, Lloyd S, Hafezparast M, Lennon-Pierce M, Eppig JT, Festing MF, Fisher EM. Genealogies of mouse inbred strains. Nat Genet 2000; 24: 23-25.

94. Ferris SD, Sage RD, Wilson AC. Evidence from mtDNA sequences that common laboratory strains of inbred mice are descended from a single female. Nature 1982; 295: 163-165.

95. Tucker PK, Lee BK, Lundrigan BL, Eicher EM. Geographic origin of the Y chromosomes in "old" inbred strains of mice. Mamm Genome 1992; 3: 254-261.

96. Bonhomme F, Guenet, J.L., Dod, B., Moriwaki, K., and Bulfield, G. . The polyphyletic origin of laboratory inbred mice and their rate of evolution. Biol J Linn Soc 1987; 30: 51-58.

97. Wade CM, Kulbokas EJ, 3rd, Kirby AW, Zody MC, Mullikin JC, Lander ES, Lindblad-Toh K, Daly MJ. The mosaic structure of variation in the laboratory mouse genome. Nature 2002; 420: 574-578.

98. Silver LM. Mouse Genetics: Concepts and applications. New York, NY: Oxford University Press; 1995.

99. Bishop CE, Boursot P, Baron B, Bonhomme F, Hatat D. Most classical Mus musculus domesticus laboratory mouse strains carry a Mus musculus musculus Y chromosome. Nature 1985; 315: 70-72.

100. Lindblad-Toh K, Winchester E, Daly MJ, Wang DG, Hirschhorn JN, Laviolette JP, Ardlie K, Reich DE, Robinson E, Sklar P, Shah N, Thomas D, Fan JB, Gingeras T, Warrington J, Patil N, Hudson TJ, Lander ES. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nat Genet 2000; 24: 381-386.

101. Wiltshire T, Pletcher MT, Batalov S, Barnes SW, Tarantino LM, Cooke MP, Wu H, Smylie K, Santrosyan A, Copeland NG, Jenkins NA, Kalush F, Mural RJ, Glynne RJ, Kay SA, Adams MD, Fletcher CF. Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse. Proc Natl Acad Sci U S A 2003; 100: 3380-3385.

102. Yalcin B, Fullerton J, Miller S, Keays DA, Brady S, Bhomra A, Jefferson A, Volpi E, Copley RR, Flint J, Mott R. Unexpected complexity in the haplotypes of commonly used inbred strains of laboratory mice. Proc Natl Acad Sci U S A 2004; 101: 9734-9739.

103. Vogel G. Genetics. Scientists dream of 1001 complex mice. Science 2003; 301: 456-457.

104. Lyon MF, Rastan S, Brown SDM. Genetic variants and strains of the laboratory mouse. New York, NY: Oxford University Press; 1996.

105. http://www.informatics.jax.org.

106. http://www.informatics.jax.org/external/festing/mouse/docs/BALB.shtml.

107. http://www.informatics.jax.org/external/festing/mouse/docs/C57BL.shtml.

108. http://www.informatics.jax.org/external/festing/mouse/docs/DDK.shtml.

109. Guenet JL, Bonhomme F. Wild mice: an ever-increasing contribution to a popular mammalian model. Trends Genet 2003; 19: 24-31.

110. Bonhomme F, Guenet JL. The laboratory mouse and its wild relatives. In: Lyon MF (ed.) Genetic Variants and Strains of the Laboratory Mouse: Oxford University Press; 1996.

111.   Chevret P. J, P., and Catzeflis, F. . Evolutionary systematics of the Indian mouse Mus famulus Bonhote, 198: molecular (DNA/DNA hybridization and 12S rRNA sequences) and morphological evidences. Zool. J. Linn. Soc. 2002; 137: 385.

112.   Lundrigan BL, Jansa SA, Tucker PK. Phylogenetic relationships in the genus mus, based on paternally, maternally, and biparentally inherited characters. Syst Biol 2002; 51: 410-431.

113.   Schalkwyk LC, Jung M, Daser A, Weiher M, Walter J, Himmelbauer H, Lehrach H. Panel of microsatellite markers for whole-genome scans and radiation hybrid mapping and a mouse family tree. Genome Res 1999; 9: 878-887.

114.   She JX, Bonhomme F, Boursot P, Thaler L, Catzeflis F. Molecular phylogenies in the genus Mus: comparative analysis of electrophoretic, scnDNA hybridization, and mtDNA RFLP data. Biol J Linn Soc 1990; 41: 83-103.

115.   http://www.informatics.jax.org/external/festing/mouse/docs/MOM.shtml.

116.   Frazer KA, Wade CM, Hinds DA, Patil N, Cox DR, Daly MJ. Segmental phylogenetic relationships of inbred mouse strains revealed by fine-scale analysis of sequence variation across 4.6 mb of mouse genome. Genome Res 2004; 14: 1493-1500.

117.   Grupe A, Germer S, Usuka J, Aud D, Belknap JK, Klein RF, Ahluwalia MK, Higuchi R, Peltz G. In silico mapping of complex disease-related traits in mice. Science 2001; 292: 1915-1918.

118.   Ebersberger I, Metzler D, Schwarz C, Paabo S. Genomewide comparison of DNA sequences between humans and chimpanzees. Am J Hum Genet 2002; 70: 1490-1497.

119.   Reich DE, Gabriel SB, Altshuler D. Quality and completeness of SNP databases. Nat Genet 2003; 33: 457-458.

120.   Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 2001; 409: 928-933.

121.   Salisbury BA, Pungliya M, Choi JY, Jiang R, Sun XJ, Stephens JC. SNP and haplotype variation in the human genome. Mutat Res 2003; 526: 53-61.

122.   Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, et al. The sequence of the human genome. Science 2001; 291: 1304-1351.

123.   Zietkiewicz E, Yotova V, Gehl D, Wambach T, Arrieta I, Batzer M, Cole DE, Hechtman P, Kaplan F, Modiano D, Moisan JP, Michalski R, Labuda D. Haplotypes in the dystrophin DNA segment point to a mosaic origin of modern human diversity. Am J Hum Genet 2003; 73: 994-1015.

124.   Kaessmann H, Wiebe V, Paabo S. Extensive nuclear DNA sequence diversity among chimpanzees. Science 1999; 286: 1159-1162.

125.   Chen FC, Li WH. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am J Hum Genet 2001; 68: 444-456.

126.   Sakate R, Osada N, Hida M, Sugano S, Hayasaka I, Shimohira N, Yanagi S, Suto Y, Hashimoto K, Hirai M. Analysis of 5'-end sequences of chimpanzee cDNAs. Genome Res 2003; 13: 1022-1026.

127.   Auffray J-C, Vanlerberghe, F., and Britton-Davidian, J. . The house mouse progression in Eurasia: a palaeontological and archaeozoological approach. Biol. J. Linn. Soc 1990; 41: 13-25.

128.   Din W, Anad, R., Boursot, P., Darviche, D., Dod, B., Jouvin-Marche, E., Orth, A., Talwar, G.P., Cazenave, P.A. and Bonhomme, F. . Origin and radiation fo the house mouse: clues from nuclear genes. J Evol Biol 1996; 9: 519-539.

129. Prager EM, Orrego C, Sage RD. Genetic variation and phylogeography of central Asian and other house mice, including a major new mitochondrial lineage in Yemen. Genetics 1998; 150: 835-861.

130. http://www.jax.org.

131. Greene-Till R, Zhao Y, Hardies SC. Gene flow of unique sequences between Mus musculus domesticus and Mus spretus. Mamm Genome 2000; 11: 225-230.

132. Orth A, Belkhir K, Britton-Davidian J, Boursot P, Benazzou T, Bonhomme F. [Natural hybridization between 2 sympatric species of mice, Mus musculus domesticus L. and Mus spretus Lataste]. C R Biol 2002; 325: 89-97.

133. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, et al. Initial sequencing and comparative analysis of the mouse genome. Nature 2002; 420: 520-562.

134. de La Casa-Esperon E, Loredo-Osti JC, Pardo-Manuel de Villena F, Briscoe TL, Malette JM, Vaughan JE, Morgan K, Sapienza C. X chromosome effect on maternal recombination and meiotic drive in the mouse. Genetics 2002; 161: 1651-1659.

135. Geserick P, Kaiser F, Klemm U, Kaufmann SH, Zerrahn J. Modulation of T cell development and activation by novel members of the Schlafen (slfn) gene family harbouring an RNA helicase-like motif. Int Immunol 2004; 16: 1535-1548.

136. Schwarz DA, Katayama CD, Hedrick SM. Schlafen, a new family of growth regulatory genes that affect thymocyte development. Immunity 1998; 9: 657-668.

137.   Babcock CS, Anderson WW. Molecular evolution of the Sex-Ratio inversion complex in Drosophila pseudoobscura: analysis of the Esterase-5 gene region. Mol Biol Evol 1996; 13: 297-308.

138.   http://genome.ucsc.edu.

139.   http://www.informatics.jax.org/external/festing/mouse/docs/RBA.shtml.

140.   Boue A, Boue J, Gropp A. Cytogenetics of pregnancy wastage. Adv Hum Genet 1985; 14: 1-57.

141.   Gropp A, Winking H. Robertsonian translocations: cytology, meiosis, segregation patterns and biological consequences of heterozygosity. Symp Zool Soc Lond 1981; 47: 141-181.

142.   Neri G, Serra A, Campana M, Tedeschi B. Reproductive risks for translocation carriers: cytogenetic study and analysis of pregnancy outcome in 58 families. Am J Med Genet 1983; 16: 535-561.

143.   Schultz R, Underkoffler LA, Collins JN, Oakey RJ. Nondisjunction and transmission ratio distortion of chromosome 2 in a (2.8) Robertsonian translocation mouse strain. Mamm Genome 2005; 17: 239-247.

144.   Gao S, Czirr E, Chung YG, Han Z, Latham KE. Genetic variation in oocyte phenotype revealed through parthenogenesis and cloning: correlation with differences in pronuclear epigenetic modification. Biol Reprod 2004; 70: 1162-1170.

145.   Latham KE. Epigenetic modification and imprinting of the mammalian genome during development. Curr Top Dev Biol 1999; 43: 1-49.

146.   Latham KE. Stage-specific and cell type-specific aspects of genomic imprinting effects in mammals. Differentiation 1995; 59: 269-282.

147.   Latham KE, Solter D. Effect of egg composition on the developmental capacity of androgenetic mouse embryos. Development 1991; 113: 561-568.

148.   Roemer I, Reik W, Dean W, Klose J. Epigenetic inheritance in the mouse. Curr Biol 1997; 7: 277-280.

149.   Surani MA, Kothary R, Allen ND, Singh PB, Fundele R, Ferguson-Smith AC, Barton SC. Genome imprinting and development in the mouse. Dev Suppl 1990: 89-98.

150.   Yang H, Bell TA, Churchill GA, Pardo-Manual de Villena F. On the subspecific origin of the laboratory mouse. Nat Genet Under review.

151.   Zhang J, Hunter KW, Gandolph M, Rowe WL, Finney RP, Kelley JM, Edmonson M, Buetow KH. A high-resolution multistrain haplotype analysis of laboratory mouse genome reveals three distinctive genetic variation patterns. Genome Res 2005; 15: 241-249.

152.   http://www.niehs.nih.gov/crg/cprc.htm.

153.   http://mouse.perlegen.com/mouse/index.html.

154.   http://www.genomedynamics.org.

155.   Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, Bleich A, Bogue M, Broman KW, Buck KJ, Buckler E, Burmeister M, Chesler EJ, Cheverud JM, Clapcote S, Cook MN, Cox RD, Crabbe JC, Crusio WE, Darvasi A, Deschepper CF, Doerge RW, Farber CR, Forejt J, Gaile D, Garlow SJ, Geiger H, Gershenfeld H, Gordon T, Gu J, Gu W, de Haan G, Hayes NL, Heller C, Himmelbauer H, Hitzemann R, Hunter K, Hsu HC, Iraqi FA, Ivandic B, Jacob HJ, Jansen RC, Jepsen KJ, Johnson DK, Johnson TE, Kempermann G, Kendziorski C, Kotb M, Kooy RF, Llamas B, Lammert F, Lassalle JM, Lowenstein PR, Lu L, Lusis A, Manly KF, Marcucio R, Matthews D, Medrano JF, Miller DR, Mittleman G, Mock BA, Mogil JS, Montagutelli X, Morahan G, Morris DG, Mott R, Nadeau JH, Nagase H, Nowakowski RS, O'Hara BF, Osadchuk AV, Page GP, Paigen B, Paigen K, Palmer AA, Pan HJ, Peltonen-Palotie L, Peirce J, Pomp D, Pravenec M, Prows DR, Qi Z, Reeves RH, Roder J, Rosen GD, Schadt EE, Schalkwyk LC, Seltzer Z, Shimomura K, Shou S, Sillanpaa MJ, Siracusa LD, Snoeck HW, Spearow JL, Svenson K, et al. The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet 2004; 36: 1133-1137.

156.   http://www.jax.org.

157.   Hara K, Nakayama KI, Nakayama K. Geminin is essential for the development of preimplantation mouse embryos. Genes Cells 2006; 11: 1281-1293.

158.   Wohlschlegel JA, Dwyer BT, Dhar SK, Cvetic C, Walter JC, Dutta A. Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. Science 2000; 290: 2309-2312.

159.   Lutzmann M, Maiorano D, Mechali M. A Cdt1-geminin complex licenses chromatin for DNA replication and prevents rereplication during S phase in Xenopus. Embo J 2006; 25: 5764-5774.

160.   Montanari M, Macaluso M, Cittadini A, Giordano A. Role of geminin: from normal control of DNA replication to cancer formation and progression? Cell Death Differ 2006; 13: 1052-1056.

161.   Nakuci E, Xu M, Pujana MA, Valls J, Elshamy WM. Geminin is bound to chromatin in G2/M phase to promote proper cytokinesis. Int J Biochem Cell Biol 2006; 38: 1207-1220.

162.   Tachibana KE, Gonzalez MA, Coleman N. Cell-cycle-dependent regulation of DNA replication and its relevance to cancer pathology. J Pathol 2005; 205: 123-129.

163.   Tachibana KE, Gonzalez MA, Guarguaglini G, Nigg EA, Laskey RA. Depletion of licensing inhibitor geminin causes centrosome overduplication and mitotic defects. EMBO Rep 2005; 6: 1052-1057.

164.   Tachibana KE, Nigg EA. Geminin regulates multiple steps of the chromosome inheritance cycle. Cell Cycle 2006; 5: 151-154.

165.   Tada S, Li A, Maiorano D, Mechali M, Blow JJ. Repression of origin assembly in metaphase depends on inhibition of RLF-B/Cdt1 by geminin. Nat Cell Biol 2001; 3: 107-113.

166.   Brady G, Boggan L, Bowie A, O'Neill LA. Schlafen-1 causes a cell cycle arrest by inhibiting induction of cyclin D1. J Biol Chem 2005; 280: 30723-30734.

167.   http://may2005.archive.ensembl.org/Mus_musculus/index.html.

168.   http://www.celera.com.

169.   http://www.idtdna.com.

170.	http://www.celera.com.

171.	Dietrich WF, Miller J, Steen R, Merchant MA, Damron-Boles D, Husain Z, Dredge R, Daly MJ, Ingalls KA, O'Connor TJ. A comprehensive genetic map of the mouse genome. Nature 1996; 380: 149-152.

172.	Pardo-Manuel de Villena F, Slamka C, Fonseca M, Naumova AK, Paquette J, Pannunzio P, Smith M, Verner A, Morgan K, Sapienza C. Transmission-ratio distortion through F1 females at chromosome 11 loci linked to Om in the mouse DDK syndrome. Genetics 1996; 142: 1299-1304.

173.	http://arbl.cvmbs.colostate.edu/molkit/dnadot/index.html.

174.	http://www.ensembl.org/Rattus_norvegicus/index.html.

175.	http://genome.ucsc.edu.

176.	http://www.ncbi.nlm.nih.gov.

177.	Felsenstein J. PHYLIP-Phylogeny inference package(version3.2). Cladistics 1989; 5: 164-166.

178.	http://www.resgen.com.

179.	http://www.sigmaaldrich.com.

180.	http://www.sequenom.com.

181.	Moran JL, Bolton AD, Tran PV, Brown A, Dwyer ND, Manning DK, Bjork BC, Li C, Montgomery K, Siepka SM, Vitaterna MH, Takahashi JS, Wiltshire T, Kwiatkowski DJ, Kucherlapati R, Beier DR. Utilization of a whole genome SNP panel for efficient genetic mapping in the mouse. Genome Res 2006; 16: 436-440.

182.	http://www.illumina.com.

183.	http://www.informatics.jax.org.

184.   Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. Bioinformatics 2003; 19: 889-890.

185.   http://bioweb.pasteur.fr/seqanal/interfaces/sreformat.html.

186.   http://www.ncbi.nlm.nih.gov.