# *Chapter 3:*

## *Refinement of the bv critical region using the existing backcross*

# *CHAPTER 3*

## REFINEMENT OF THE *BV* CRITICAL REGION USING THE EXISTING BACKCROSS

### 3.1 INTRODUCTION

This chapter describes the work carried out to characterise and refine the candidate region for the *bronx waltzer* gene as defined by a pre-existing intraspecific backcross of [(*bv*/*bv* x 101/H)F1 x *bv*/*bv*]. It also includes the determination for the first time of the physical size of the critical interval.

#### 3.1.1 Genetic mapping

Recombination events which occur between sister chromatids during meiosis are the phenomena that make genetic mapping possible. Genetic mapping determines the linkage between two given markers based on the frequency with which they recombine. The closer they are to each other, the lower the probability that a crossover will occur between them. Hence a marker which is seen to co-segregate closely with a phenotype can be an indication that the allele responsible for the phenotype is located close to the marker. The second requirement for genetic mapping is a means of distinguishing between copies of a marker inherited from each parent. The markers used must be polymorphic – be able to exist as multiple variations within a given population. In human genetic mapping this is achieved by assessing the haplotypes of different family members in order to establish which portions of the genome have been inherited from each parent. In the mouse, it is made simpler by the ability to control matings and so introduce polymorphisms into a population by crossing the mutant mouse with a wild type mouse carrying different alleles of the genetic markers to be used. The most common strategy, and the one used here, is the backcross.

#### 3.1.1.1 Establishing a backcross

The first stage of the backcross is to mate the mutant strain with a different strain, known as an outcross. In this case the outcross was between *bronx waltzer* (*bv*/*bv*) and the inbred strain 101/H (+/+ at the *bronx waltzer* locus). The resulting $F_1$ offspring will all be +/*bv*, carrying one chromosome from each parent. Since the *bronx waltzer* mutation is recessive, these are then mated with homozygous *bv*/*bv* mice so as to give backcross progeny which may be either +/*bv* or *bv*/*bv* and allow phenotyping according to their behaviour. Each backcross mouse will carry one chromosome purely from the *bronx waltzer* background and one from the outcross parent. This chromosome from the $F_1$ parent may have undergone recombination between *bv* and 101/H DNA, allowing the position of polymorphic markers to be mapped in relation to the phenotype of the mouse. This backcross strategy is illustrated in Figure 3.1.

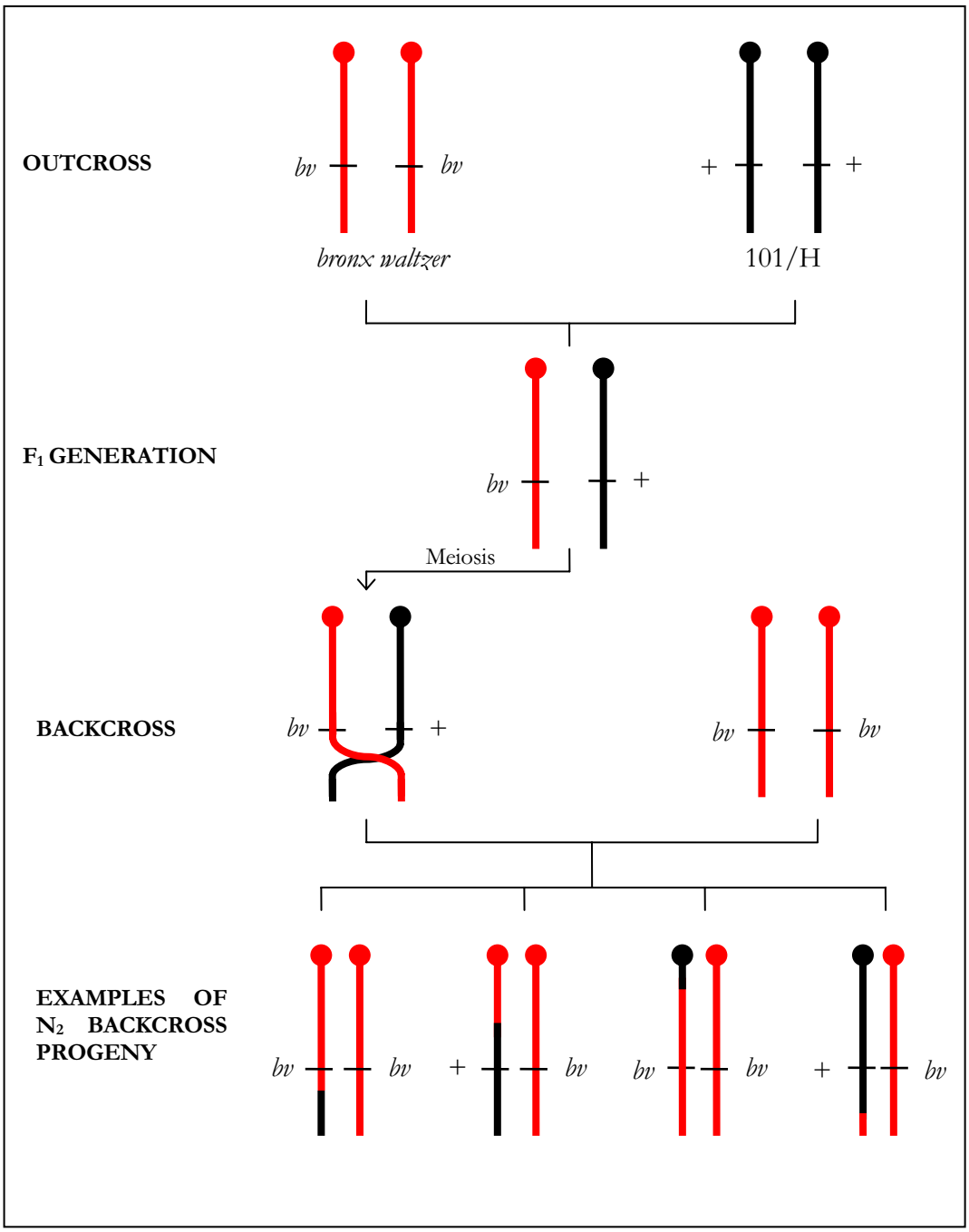



**Figure 3.1**: The backcross strategy employed in the genetic mapping of the *bronx waltzer* locus. Chromosomes derived from *bronx waltzer* are shown in red while those from the inbred strain 101/H are in black. During gamete formation in the F1 generation, recombination may occur between the two parental chromosomes at the prophase I phase of meiosis. These recombinant chromosomes are inherited by the backcross offspring along with one set of *bronx waltzer* chromosomes, giving a variety of possible progeny as shown in the bottom row.

### 3.1.1.2 Haplotype analysis

Study of the recombined chromosomes inherited from the $F_1$ generation mice by backcross offspring is known as haplotype analysis. The frequency of recombination between two markers is used to calculate the relative genetic distance between them and the *bronx waltzer* allele, as indicated by the phenotype of the mouse, can be placed amongst them. Relative genetic distances are expressed as centiMorgans (cM), where two loci are said to be one centiMorgan apart if recombination is observed between them in 1% of meioses.

### 3.1.1.3 Markers for genetic mapping

In order to be informative within the context of a genetic mapping approach, markers must exhibit discernable differences between the genetic backgrounds being utilised, a property known as polymorphism. Polymorphic markers allow the parental origin of each portion of an individual's genetic make-up to be determined, thus enabling the linkage of different portions of the genome to the phenotypic trait in question to be traced.

The simplest form of polymorphism to detect is one where a stretch of sequence varies in length between strains. This is generally due to the presence of a repeat sequence which is non-coding and can exist in varying copy numbers (Hamada *et al.* 1982). Such sequences, often referred to as microsatellites or simple sequence length polymorphisms (SSLPs), are di-, tri-, or tetra-nucleotide repeats which exist at high frequency within the genome. Their varying sizes between different individuals allow simple analysis by PCR using primers designed to sequence flanking the repeat region, making them very useful as genetic markers. An example of an SSLP is shown in Figure 3.2.

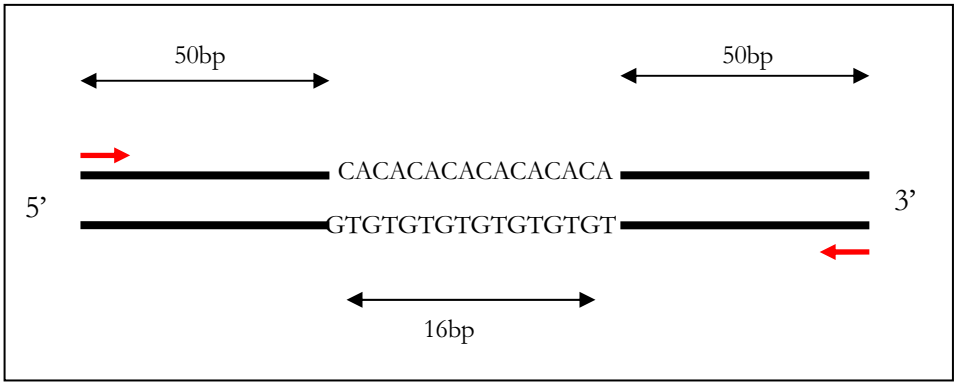



**Figure 3.2**: Diagram showing a typical simple sequence length polymorphism (SSLP). The example given is a di-nucleotide repeat. The number of CA units varies between individuals, giving differently sized fragments when the region is amplified by PCR using the primers shown in red.

A genetic linkage map placing 6580 such markers to a resolution of 0.2cM within the mouse genome was constructed by Dietrich *et al.* (1996). This framework map of markers greatly simplifies the process of identifying a candidate region for a given gene, since backcross mice can be quickly screened with a panel of markers chosen to cover the whole genome to initially locate a region of linkage. Markers can then be chosen at higher density from within that region in order to narrow the area of interest.

A second form of polymorphism which can be useful for genetic mapping is the single nucleotide polymorphism (SNP). These single base substitutions are present at varying densities throughout the mouse genome and may be found in both coding and non-coding regions. The more distantly two strains of mice are related, the more SNPs with differing alleles can be expected to be found. In general SNP discovery is achieved by sequence analysis, and once detected large numbers of samples can be screened for the different alleles in a number of ways. If the base change occurs within the recognition site of a restriction enzyme and disrupts its ability to bind, it is possible to rapidly assay it by performing a restriction digest and analysing the pattern of fragment sizes obtained. An alternate method of SNP screening is to perform an Oligo Ligation Assay (OLA), which requires a perfect sequence match between a pair of oligos and the target sequence in the amplified fragment to give a

band. If the polymorphic base is present at the point where the two oligos touch, the oligos can be ligated together and produce a band of the expected size. Additionally, the assay can be designed to give various different sized bands by using several sets of oligo probes, and different fluorescent tag colours can also be incorporated.

#### 3.1.1.4 Phenotypic analysis

Mice were typed phenotypically at the *bronx waltzer* allele by assessment of their behaviour. Two behavioural parameters were analysed – the presence or absence of the Preyer reflex as an indicator of basic auditory function and manifestation of shaker-waltzer behaviour as a measure of vestibular dysfunction. Prior to sacrificing the backcross offspring for DNA preparation, these tests were performed to determine the phenotype of each mouse and so to establish which copy of the *bv* allele had been inherited from the $F_1$ parent. Since all mice inherit one mutant copy of the *bv* gene from the backcross parent, manifestation of the *bronx waltzer* phenotype can be interpreted to mean that the mouse must have inherited a second mutant copy from the $F_1$ parent. Conversely, mice with normal behaviour and responses must have inherited a wild type copy of the gene, making them heterozygous at the locus.

In the course of this backcross the Preyer reflex test was administered by striking a large pair of metal forceps on a metal surface out of sight and away from the cage so as to minimize vibration. A positive result was recorded if the mouse retracted their pinnae in response to this broad frequency band of sound. In the absence of this response a negative result was recorded. Shaker-waltzer behaviour was assessed by observing the mice and a positive result was recorded if the mouse demonstrated rapid upward flicking motions of the head and periodic circling within the cage. In cases where the phenotype was questionable, a reaching response test was administered. Here, the mouse was held by the tail above the cage and its behaviour observed. Mice exhibiting

shaker-waltzer behaviour curl their heads towards their belly and/or revolve around their body axis while normal mice stretch their limbs and reach out towards the ground.

Mice thought to be homozygous for *bv* were removed from the cage first in order to prevent their hyperactive behaviour from disturbing their heterozygote littermates and causing them to behave atypically as a result of nervousness. The backcross progeny were all culled by cervical dislocation and their tissue taken and stored as described in Section 2.2.

### 3.1.2 The existing *bv*/101 genetic map

The backcross between *bronx waltzer* and 101/H was originally established by Bussoli (1996). The inbred strain 101/H was chosen as the outcross background following analysis of 33 SSLP markers mapping to mouse chromosome 5, to which the mutation had previously been localised (R. Brister, *pers. comm.*). 101/H showed 52% polymorphism with *bronx waltzer*, making it the most polymorphic of the inbred strains tested. In an initial typing of 701 backcross mice, the mutation was mapped to a 1.28cM region between the markers *D5Mit25* and *D5Mit188* (Bussoli *et al.* 1997).

Following subsequent breeding and work by Cheong (2000), a final total of 1085 backcross mice arose from this cross. Twelve of these mice were excluded from further analysis as a result of either discrepant or difficult phenotyping, leaving 1073 for haplotype analysis. During the course of this study, the single mouse (T672) which had positioned the *bv* locus distal of the marker *D5Mit209* was retyped and its breakpoint found to lie instead between *D5Mit367* and *D5Mit95*. As a result, and with the inclusion of data arising from the new backcross mice, the new candidate region for *bronx waltzer* was described as a 1.86cM region flanked by the markers *D5Mit25* and *D5Mit209*. The haplotype diagram representing these results is shown in Figure 3.3, and the genetic map arising from these data is shown in Figure 3.4.

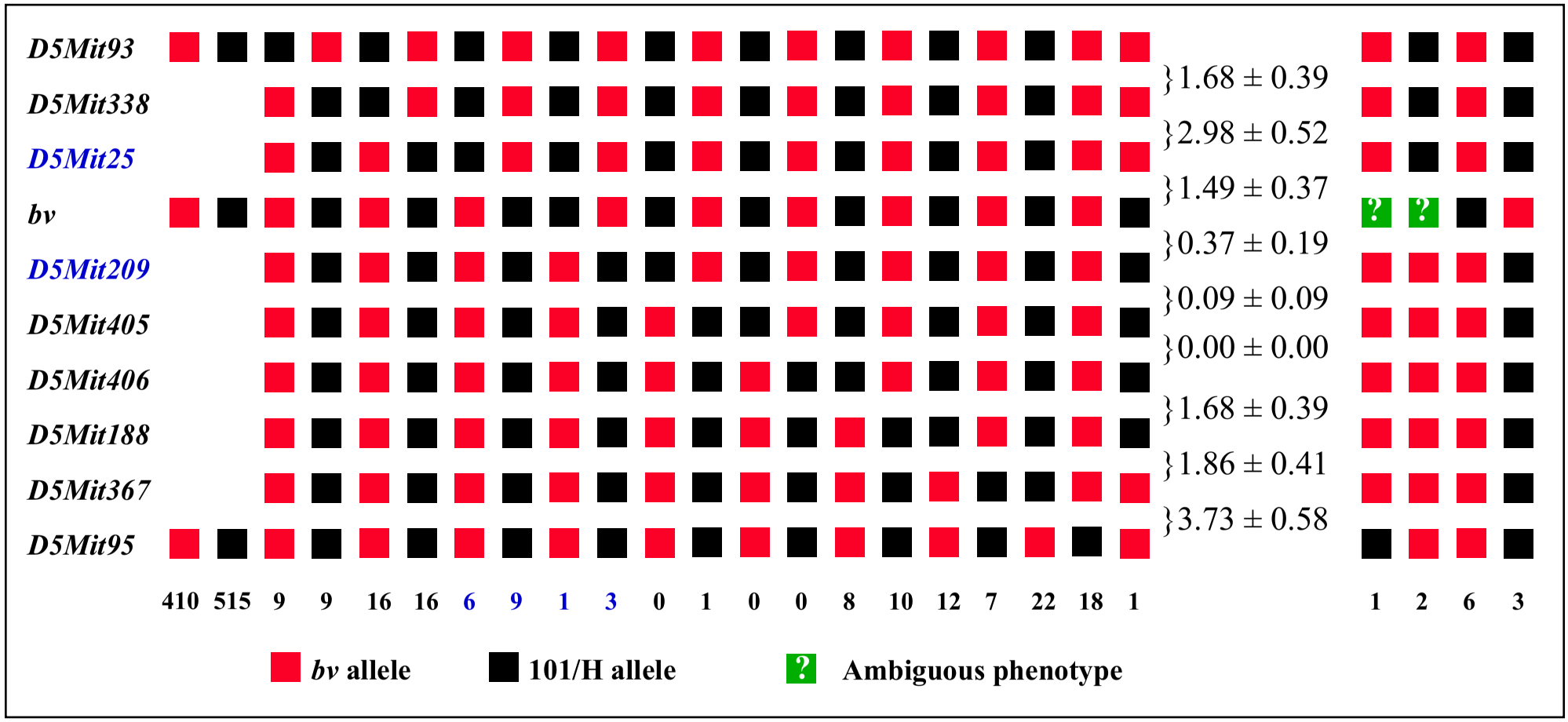



**Figure 3.3:** Haplotype analysis of the total 1085 progeny arising from the *bv*/101/H intraspecific backcross. Genetic markers used in the mapping are listed on the left, and numbers of mice typed with each pattern of recombination are shown along the bottom. Genetic distances calculated between the markers based on the number of recombinations occurring between them are given to the right. Mice excluded from the analysis as a result of discrepant or difficult phenotyping are shown on the far right of the diagram. Flanking markers and mice exhibiting recombinations between them are highlighted in **blue**. Adapted from Cheong (2000).

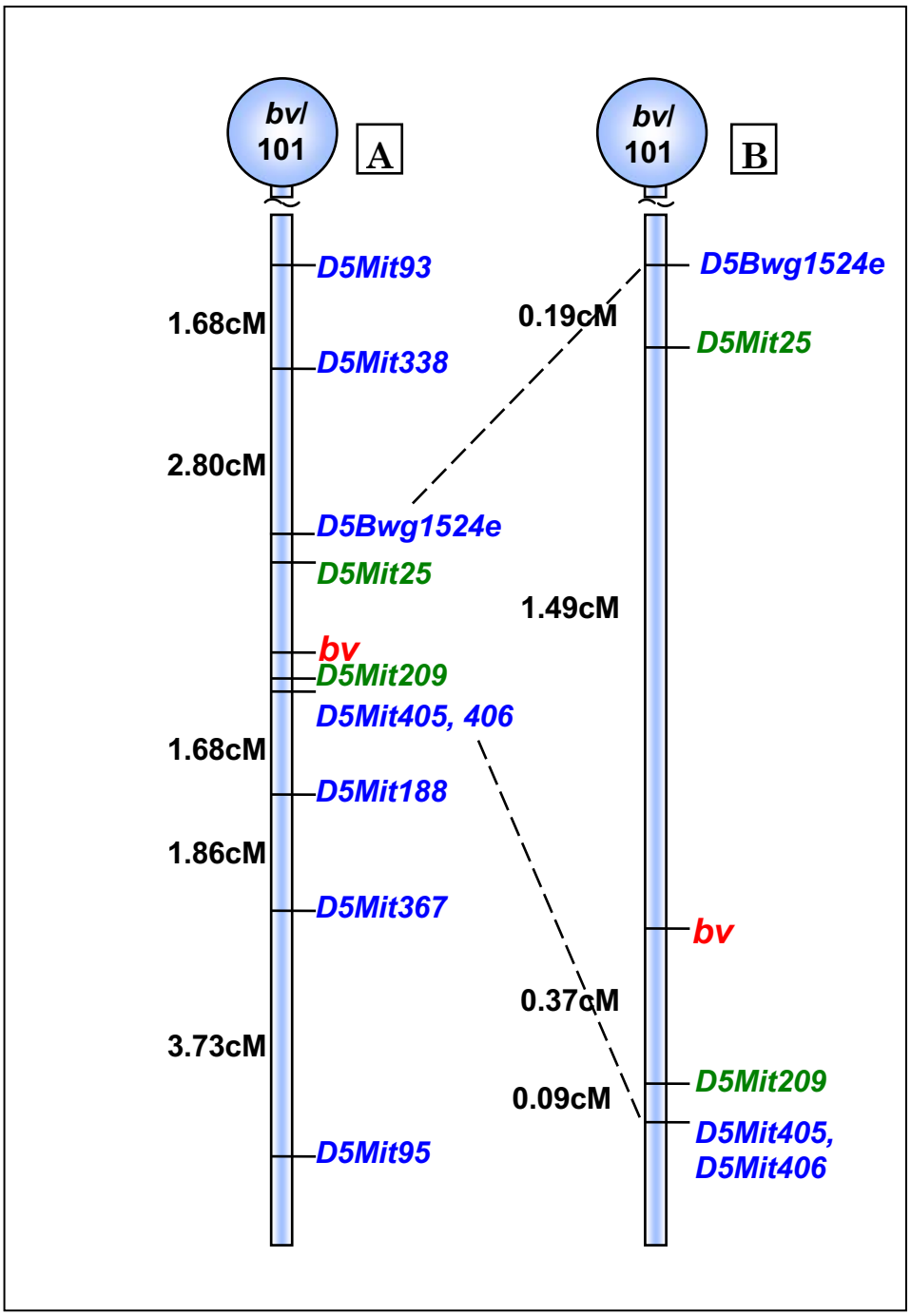



**Figure 3.4:** The mouse genetic map as elucidated using the 1073 backcross mice typed from the *bv*/101 cross. [A] depicts the region on chromosome 5 surrounding the *bv* locus, the markers closest to *bv* are shown on an enlarged map in [B].Genetic loci are drawn to the right of the diagrams, with the *bronx waltzer* locus shown in red, flanking markers in green and other genetic markers in blue. Genetic distances determined by the number of recombinations observed between markers are given in black to the left of the diagrams.

### 3.1.3 Identifying new markers

As shown by the haplotype diagram in Figure 3.3, there remain 19 backcross mice exhibiting recombinations between D5Mit25 and D5Mit209. Therefore this backcross still holds the potential to narrow the *bv* candidate region further if novel polymorphic markers can be found which map between the flanking markers. Previous attempts made at identifying such markers have included the screening of the European Collaborative Interspecific Backcross (EUCIB) mapping panel and the Jackson Laboratory mapping panels BSB and BSS in order to identify ESTs or genes which may include polymorphisms (Cheong 2000). This work also helped to identify a region of conserved synteny on human chromosome 12, from which gene sequence was obtained. One of the markers identified in this manner, D5Bwg1524e, proved to be polymorphic for *bv* and 101, but on screening the backcross panel it was found to map proximal to D5Mit25 and therefore lies outside the critical region as shown in Figure 3.4.

The paucity of markers in the region and the potential remaining in the backcross are addressed in this chapter with a continuation of the search for new markers using resources and sequence that were only made available more recently.

### 3.1.4 Previous physical mapping

Using the genetic map determined by the *bv*/101 backcross, the next stage of positional cloning began with the construction of a physical map to cover the critical interval (Cheong 2000). The genetic markers which mapped close to the *bronx waltzer* locus were first used to screen a YAC library but the clones identified in this way were suspected of harbouring deletions and chimerisms and further work was abandoned. The second library screen was of the CITB Mouse BAC library but this too was problematic and resulted in a large number of false positives. A map was built using the data obtained and the distances between markers estimated based on the sizes of the BACs for

which they were positive. Using this methodology the physical distance between the flanking markers *D5Mit25* and *D5Mit209* was determined to span under 1Mb since both markers were found to be positive for, and therefore present within a single 91Kb clone, BAC42I5. Given the 1.86cM genetic distance separating these markers, this physical distance seemed remarkably small and the results were called into question. In the mouse genome 1cM is thought to equate to roughly 2Mb of sequence so the critical interval might be expected on average to be 3.72Mb. This discrepancy could potentially be explained by the presence of a recombination hotspot located between the markers causing a higher than average recombination rate. Further physical mapping as described within this chapter allows the comparison of these data to those arising from the public mouse mapping and sequencing project.

## 3.2 METHODS

Some of the methods employed in this chapter are described in Chapter 2. The following procedures are specific to this chapter.

### 3.2.1 Bacterial Artificial Chromosomes (BACs)

The BACs used form part of the Mouse C57BL/J6 Female RPCI-23 and Mouse C57BL/J6 Male RPCI-24 BAC libraries originating from the Roswell Park Cancer Institute, Buffalo, New York. The RPCI-23 library was constructed in the pBACe3.6 vector shown in Figure 3.5 (Frengen *et al.* 1999)from *Eco*RI-digested C57BL/6J genomic DNA (Osoegawa *et al.* 2000). The RPCI-24 library is made up of size-selected *Mbo*I fragments cloned into the pTARBAC1 vector (Zeng *et al.* 2001) between the *Bam*HI sites (see Figure 3.6). Both vectors are hosted in the *Escherichia coli* substrain DH10B and include chloramphenicol resistance genes allowing for selection by growth on media containing the antibiotic. Clones originating from the RPCI-23 library are generally identified by the prefix bM, and those from the RPCI-24 library by the prefix bN.

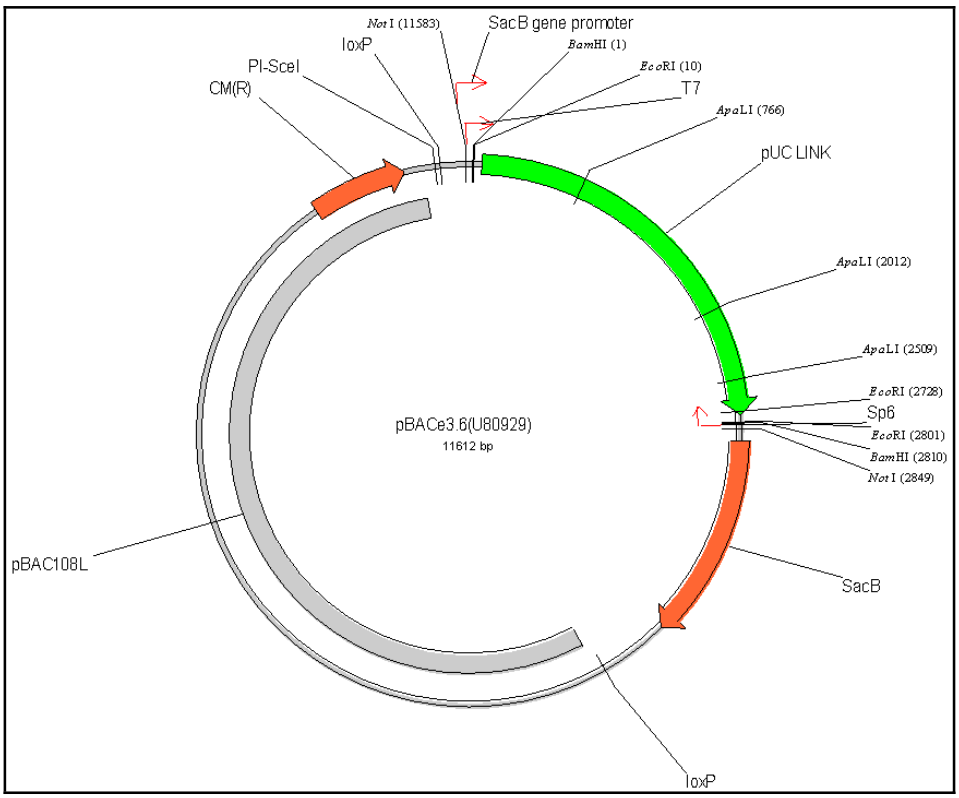



**Figure 3.5:** Diagrammatic representation of the cloning vector pBACe3.6 used in the construction of the female mouse C57BL/6J BAC library RPCI-23 (Frengen *et al.* 1999).

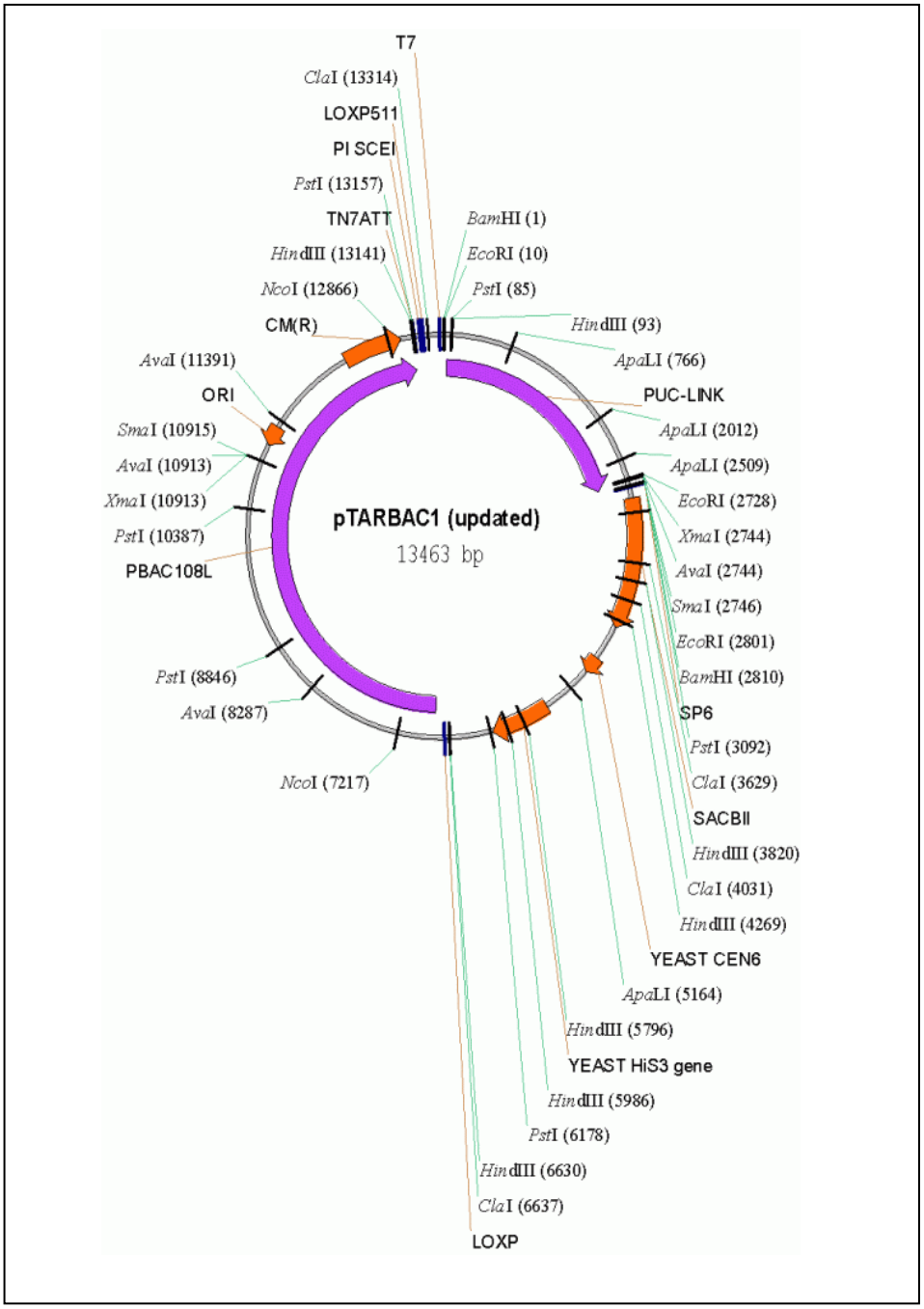



**Figure 3.6:** Diagrammatic representation of the cloning vector pTARBAC1 used in the construction of the male mouse C57BL/6J BAC library RPCI-24 (Zeng *et al.* 2001).

Clones were obtained from the BACPAC Resources Center (BPRC) at Children's Hospital Oakland Research Institute in Oakland, California and arrived as bacterial stab cultures. On arrival these were streaked onto LB agar (10 g/l bacto-tryptone; 5 g/l yeast extract; 10 g/l NaCl; 15 g/l agar; pH7.4) containing 20µg/ml chloramphenicol and incubated overnight at 37°C. A single colony from each culture was then used to inoculate 11ml LB broth (10 g/l bacto-tryptone; 5 g/l yeast extract; 10 g/l NaCl; pH7.4) containing 20µg/ml chloramphenicol and incubated overnight at 37°C with vigorous shaking (300rpm). These cultures were then used for the preparation of plasmid DNA as well as for long-term storage of the clones.

#### 3.2.1.1 Storage of BAC clones

From each LB culture grown from a single bacterial colony, two aliquots of 400µl were removed to two separate 1.5ml microcentrifuge tubes labelled with the name of the clone from which they were derived. To each of these was added 100µl sterile 50% glycerol, giving a final glycerol concentration of 10%. These were then stored separately at -80°C, with one set of tubes being used to inoculate subsequent cultures and one set being reserved as archive copies.

#### 3.2.1.2 Preparation of plasmid DNA from BAC clones

Plasmid DNA was obtained by alkaline lysis of the bacterial culture prepared as described in Section 3.2.1. The remaining 10ml culture was centrifuged at 12,000g and 4°C for 10 minutes and the supernatant removed and discarded. The cell pellet was then resuspended in 200µl Solution I (50mM glucose; 25mM Tris-HCl, pH8; 1mM EDTA ), transferred to a 1.5ml microcentrifuge tube and incubated on ice for 5 minutes. Cells were lysed by the addition of 400µl briefly cooled Solution II (0.2M NaOH; 1% SDS) and mixed by inversion before being incubated on ice for a further 10 minutes. Precipitation of cellular debris and genomic DNA was achieved by the addition of 300µl

ice cold solution III (3M potassium acetate; 11.5% glacial acetic acid, pH4.8), mixing by inversion and centrifugation at 12,000g and 4°C for 10 minutes. The supernatant was then removed to a fresh tube to which was added 600µl ice cold isopropanol in order to precipitate plasmid DNA. Samples were incubated on ice for 10 minutes before being centrifuged at 12,000g and 4°C for 15 minutes. The supernatant was then discarded and the pellet resuspended in 200µl T0.1E (10mM Tris-HCL; 0.1mM EDTA). Plasmid DNA was purified by the addition of 200µl 50:50 phenol:chloroform, vigorous shaking and centrifugation at 12,000g and 4°C for 5 minutes. The aqueous layer was removed to a fresh tube and DNA re-precipitated by the addition of 20µl 3M sodium acetate and 200µl isopropanol. Samples were then incubated on ice for 10 minutes before being centrifuged at 12,000g and 4°C for 15 minutes. The supernatant was removed and the pellet washed by the addition of 200µl 70% ethanol followed by centrifugation at 12,000g and 4°C for 5 minutes. Following removal of the supernatant, pellets were allowed to air dry completely before being resuspended in 50µl T0.1E (10mM Tris-HCL; 0.1mM EDTA). RNA was removed by the addition of 1µl 10mg/ml RNase and incubation at 37°C for 1 hour. DNA was then quantified using a UV spectrophotometer (see Section 2.3) and dilutions made for use as template DNA in subsequent PCR reactions. Remaining DNA was stored at -20°C.

#### 3.2.1.3 Building a BAC contig to span the region

The following work was carried out prior to the publication of the map of the mouse genome (Gregory *et al.* 2002) but was enabled by access to the raw mapping data during assembly by the authors. These data consisted of clones from the BAC libraries RPCI23 and RPCI24 (see Section 3.2.1) positioned primarily using *Hin*dIII fingerprints of the clone inserts (Marra *et al.* 1997; Soderlund *et al.* 2000) alongside comparison of BAC end sequences (Zhao *et al.* 2001) to the tile path of sequenced human clones. These together were used to assemble the clones into contiguous series where neighbouring clones

shared significant restriction fragment lengths on digestion with *Hin*dIII, and where the order of BAC end sequences matched that found in the human sequence. This process of contiguation was carried out using the software package FPC (Soderlund *et al.* 2000) which is currently available from the Arizona Genomics Institute website (http://www.genome.arizona.edu/fpc/). Markers were added to the map by electronic PCR (Schuler 1997) or by hybridisation with overgo probes (Han *et al.* 2000).

The flanking markers *D5Mit25* and *D5Mit209* were localised to the pre-publication map by sequence identity. Using the raw mapping data, a tile path of 32 BACs was selected which, judging from their restriction fingerprints, would represent the complete region between the flanking markers. This was achieved by ensuring that adjacent tile path clones shared a number of restriction fragments of the same length.

### 3.2.2 Investigating potential new polymorphisms

A number of different approaches were used in order to garner information regarding new regions of polymorphism between the *bv* stock genetic background and that of the chosen backcross strain 101/H.

#### 3.2.2.1 Mouse SNP Database

The first of these was the employment of the Mouse SNP Database (http://mousesnp.roche.com/cgi-bin/msnp.pl) – a database holding records of reported single nucleotide polymorphisms between inbred strains which predated the more extensive SNP data now available through Mouse Ensembl (http://www.ensembl.org/Mus_musculus/). This database is searchable using the genetic distances determined by the MIT F2 Intercross (Dietrich *et al.* 1996) and available on the Whitehead Institute/MIT Center for Genome Research website (http://www.broad.mit.edu/cgi-bin/mouse/sts_info). The positions of the flanking markers *D5Mit25* and *D5Mit209* are 49.5cM and 49.2cM respectively, and these values were used to

obtain records from within the database which may lie within the critical region. Because the *bronx waltzer* mutation is not maintained on in inbred strain background the database could not be used to determine whether the SNPs would be polymorphic within the backcross samples, but was instead used as an indicator of potential polymorphism to be determined by PCR (see Section 2.4) or by sequencing (see Section 2.8). The SNPs obtained from this database and the sequences of primers used to screen for them are given in Appendix A.1.

#### 3.2.2.2 Sequence sampling

The second approach used was one of sequence sampling within the critical region. This effort also predated the release of the mouse genome sequence and so was limited to those areas where sequence was known and hence primers could be designed. Sequences utilised included 20 from 3' untranslated regions (UTRs) of genes mapping genetically within the region and 43 BAC end sequences obtained using primers placed within vector sequence to amplify part of the cloned insert (Kelley *et al.* 1999; Zhao *et al.* 2001) and available via the NCBI Clone Registry (http://www.ncbi.nlm.nih.gov/genome/clone/). Where a sequence was too large to be amplified by a single set of primers, several pairs of primers were designed to amplify overlapping regions of the sequence and so cover its full length. These short sequences were screened first for size polymorphisms between *bv* and 101/H by PCR and agarose gel electrophoresis and later for smaller polymorphisms by sequencing. The primers used to amplify 3' UTRs are in Appendix A.2, and those used to assay BAC ends are given in Appendix A.3.

#### 3.2.2.3 Tandem repeats

With the publication of the first draft of the mouse genome (Gregory *et al.* 2002), new methods became available in the search for novel polymorphisms.

The third approach used an algorithm designed to find regions of tandem repeats within a specified sequence (Benson, 1999), in the hope that some of these would prove to be SSLPs. Software based on this algorithm was obtained from the Boston University website (http://tandem.bu.edu/trf/trf.html) and sequence lying between the flanking markers was processed using this program. Regions showing tandem repeats of large size and high repeat conservation were selected as being most likely to be polymorphic. Of these, 18 were chosen which were distributed throughout the candidate region. Primers were designed around each tandem repeat (see Appendix A.4) and used to amplify fragments from *bv* and 101/H templates which could then be compared for size differences using agarose gel electrophoresis. Later, they were also sequenced in an effort to detect smaller size differences which may not be resolved by electrophoresis.

#### 3.2.2.4 SNPs from published sequence

A growing number of SNPs from various strains in comparison to the published C57BL/6J sequence have been reported by researchers and collated by the NCBI in dbSNP (http://www.ncbi.nlm.nih.gov/SNP/index.html). These have also been annotated onto the published sequence and can be accessed using Mouse Ensembl (http://www.ensembl.org/Mus_musculus/). Once again, primers were designed in the flanking sequence of 70 such SNPs which lay within the candidate region (see Appendix A.5 for primer sequences) and used to amplify fragments from *bv* and 101 templates by PCR, followed by sequencing of the products and comparative analysis to assess their polymorphism.

Recently, a new resource of potential SNPs became available as a result of the comparison of sequence from the strain 129S5/SvEv$^{Brd}$ to the published sequence for C57BL/6J. This sequence sampling of 129S5/SvEv$^{Brd}$ was a result of the Mutagenic Insertion and Chromosome Engineering Resource (MICER) (Adams *et al.* 2004), a publicly available library of insertional

targeting vectors. In order to determine the position of the clones, their ends were sequenced and compared to the published C57BL/6J sequence with an identity of over 100bp with >95% identity required to confirm a location. The differences between them were then annotated as SNPs. Primers were designed to flank 119 such SNPs which lay within the *bv* region of interest (primer sequences are listed in Appendix A.6) and used to amplify fragments from *bv* and 101 templates. The products were then sequenced in order to determine whether they were also polymorphic between these strains.

## 3.3 RESULTS

### 3.3.1 A physical map of the *bronx waltzer* candidate region

Through the selection of BAC clones based on their restriction digest fingerprints, a preliminary tiling path across the region was established. Clones which had end sequences available were prioritised in order to simplify the subsequent verification of the tile path. Overlaps were tested by PCR amplification using primers designed within the end sequences of a given BAC. These were used to amplify fragments from template DNA prepared from the inserts of the BACs thought to neighbour the one being tested. DNA from the two clones on each side of each BAC was used in order to check that no clones were redundant in the tile path. This strategy is illustrated in Figure 3.7 and examples of the products obtained in these amplification assays are shown in Figure 3.8.

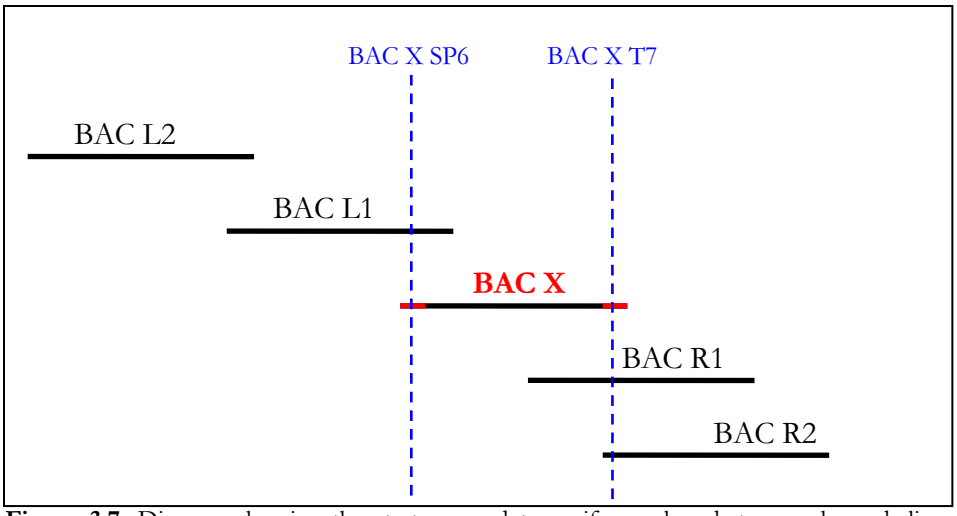



**Figure 3.7:** Diagram showing the strategy used to verify overlaps between clones believed to be contiguous. In this example the overlaps of BAC X are being verified. Red bars represent the end sequences of the BAC insert obtained by sequencing using primers designed to the flanking vector sequences, either SP6 or T7. The primers BACXSP6 and BACXT7 are designed within these sequences and used to amplify fragments from template DNA prepared from the proposed neighbouring clones. Successful amplification is taken as a positive result, meaning that the two clones overlap and shown here by blue dashed lines. In this case, BAC X overlaps BAC L1 to the left and BACs R1 and R2 to the right, making BAC R1 redundant.

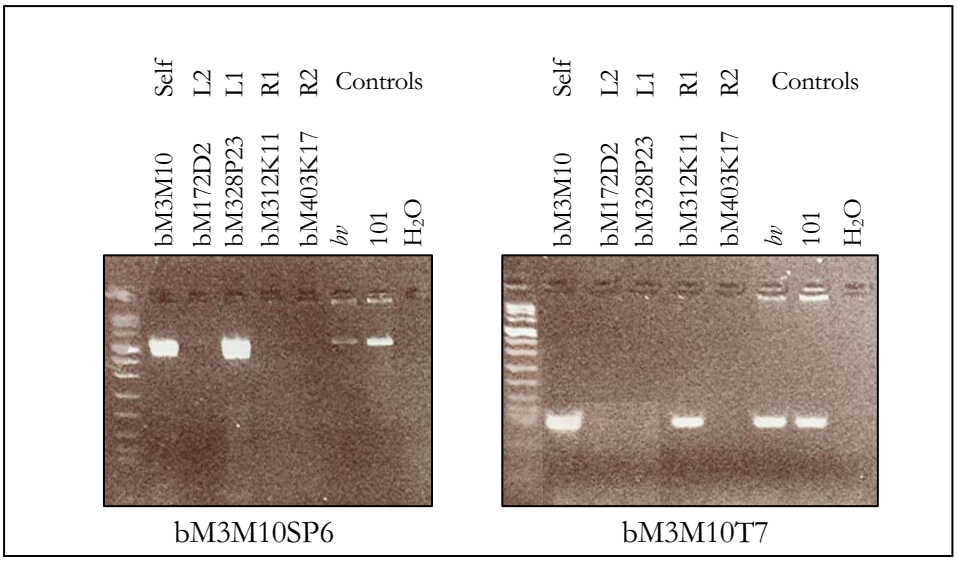



**Figure 3.8:** An example of the assay used to confirm BAC overlaps, showing the results for clone bM3M10. Primers designed to the SP6 end sequence are positive for the L1 clone bM328P23, while those designed to the T7 end sequence are positive for the R1 clone bM312K11. The L2 and R2 clones give negative results, showing that no clones in this portion of the tile path are redundant.

### 3.3.1.1 Placement of genetic mapping markers

Once the overlaps had been verified and the clones confirmed to be contiguous but non-redundant, the panel of DNA from each BAC insert was tested by PCR amplification with each of the known markers from the genetic map in order to define physical size of the critical region. Amplified products visualised by agarose gel electrophoresis were taken as evidence that a given marker was contained within the clone which gave a positive result. Some examples of this assay are shown in Figure 3.9 and the completed physical map showing confirmed overlaps and marker positions is presented in Figure 3.10.

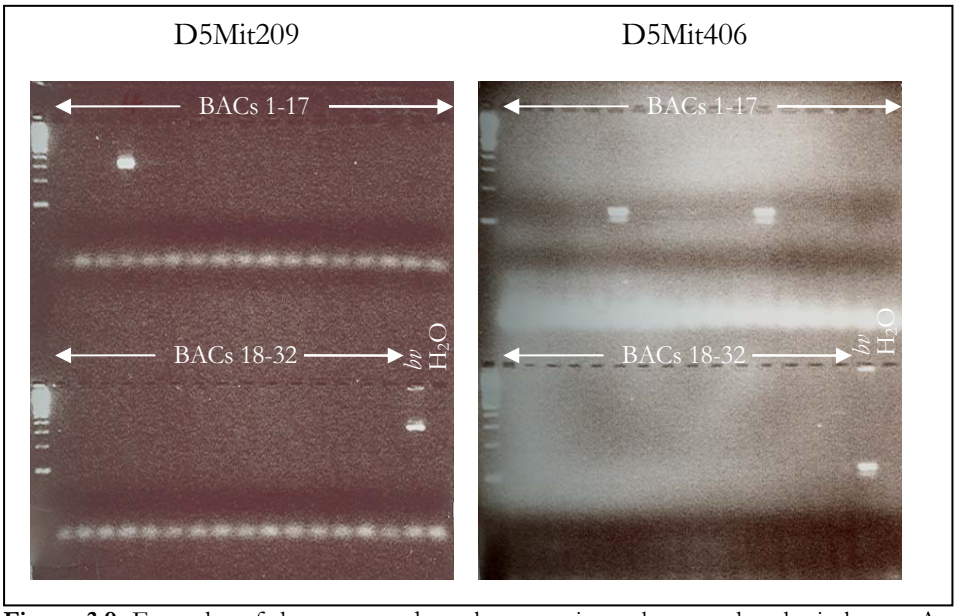



**Figure 3.9**: Examples of the assay used to place genetic markers on the physical map. A panel of DNA prepared from BAC clones in the minimum tiling path was screened by PCR using primers designed to the markers. A positive result for a clone suggests that it contains that marker. *D5Mit209* was positive for clone bM75L23, while D5Mit406 hits overlapping clones bM79M23 and bM231H24.

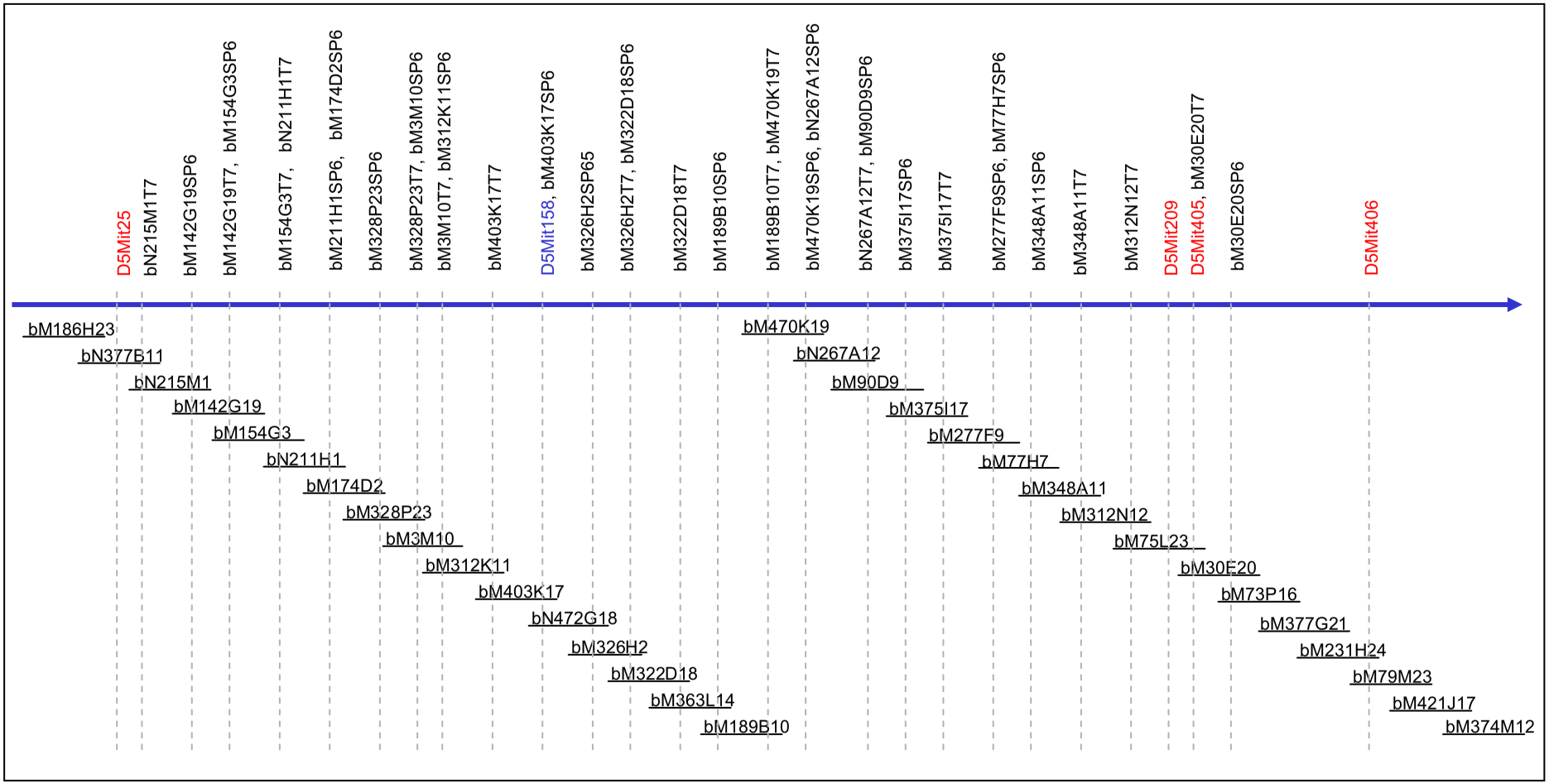



**Figure 3.10:** A contiguous BAC physical map of the *bronx waltzer* candidate region. BAC end sequences (shown in black) were used as markers to confirm overlaps, positive results are represented as dashed grey lines. Genetic markers polymorphic for *bv* and 101/H are shown in **red**, those not polymorphic are shown in **blue**.

#### 3.3.1.2 Estimating the physical size of the region

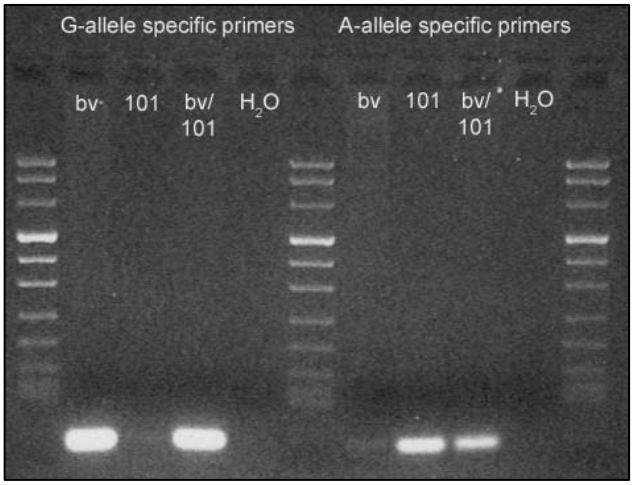
With the flanking markers shown to be contained within clones bN377B11 (*D5Mit25*) and bM75L23 (*D5Mit209*) and the mapping between them proven to be contiguous, it became possible to estimate the physical distance between them. This was initially achieved using the software FPC (see Section 3.3.1.1) which is able to estimate distances based on the cumulative size of fragments in the *Hin*dIII fingerprints of clones within a given region. The distance obtained by this method between the flanking markers *D5Mit25* and *D5Mit209* was 2.6Mb. Later, as the mouse sequence became available, the physical distance was calculated as 2.8Mb, showing the mapping estimate to be quite accurate.

### 3.3.2 Identifying new polymorphic markers

In order to be informative within the context of the existing and still powerful panel of 1073 mice from the *bv*/101 backcross, markers must be distinguishable from one background to the other so that it can be determined from which parent a mouse has inherited that marker. This section details the outcome of the approaches used to identify such markers and reduce the size of the candidate region, thus also reducing the number of genes to be considered as candidates for *bronx waltzer*.

#### 3.3.2.1 Publicly available SNPs

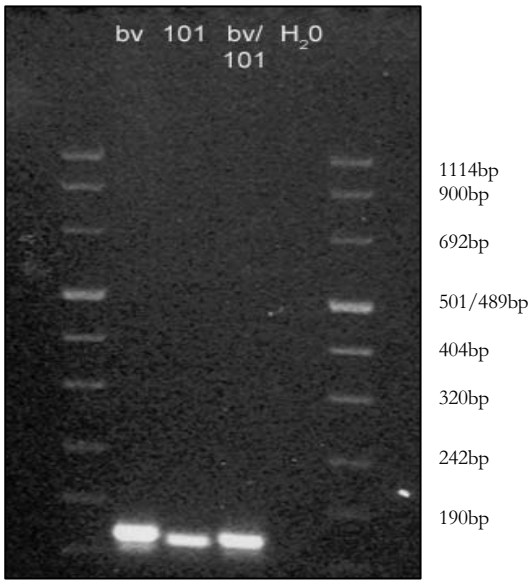
The first source of polymorphisms to be investigated was the Mouse SNP Database. Three markers were found to map close to the *bv* region and these were tested for polymorphism between *bv* and 101/H by PCR assay. Of the three markers, two were shown to be polymorphic between the two strains. M09391, a SNP represented by a G in *bv* and an A in 101/H, was detected using allele-specific PCR primers. Two reverse primers were designed to the sequence where the SNP was located, one including the G-allele and the other incorporating the A-allele. Using a common forward primer, these gave amplification only when the included allele was present, as shown in Figure 3.11. M26940, a short polymorphic repeat, was shown to be slightly longer in 101 than in *bv*, as shown in Figure 3.12. The third marker obtained from the Mouse SNP database, M11668, was tested by PCR assay but was found not to be polymorphic between the two strains used in this cross (data not shown).




**M09391**

AACACCAG**[A/G]**ACTGAAT

**Figure 3.11**: Publicly available SNP M09391, showing allele-specific amplification and demonstrating polymorphism between *bv* and 101/H. The marker lane contains Molecular Marker VIII (Roche).




**M26940**

ATCTAT**[CTAT/]**TTATTT

**Figure 3.12**: Polymorphic repeat M26940 showing differing sizes on amplification from *bv* and 101/H templates. The marker lane contains Molecular Marker VIII (Roche).

The markers obtained from the Mouse SNP database were mapped close to the *bv* locus according to genetic distances from the MIT F2 Intercross (http://www.broad.mit.edu/mouse_rh/) but their precise positions in relation to the existing flanking markers were not known. Therefore, before screening the backcross panel of recombinant mice with these newly identified markers, their position within the region was determined by screening the panel of BAC inserts which make up the physical map described in Section 3.3.1. The resulting position of the markers is shown in Figure 3.13. This screen demonstrated that the two polymorphic markers lie distal of *D5Mit209*, are outside the candidate region and therefore cannot be used to narrow the region further. The marker M11668 does lie within the critical region but cannot be used for genetic mapping purposes because *bv* and 101 alleles do not demonstrate polymorphism and therefore cannot be differentiated.

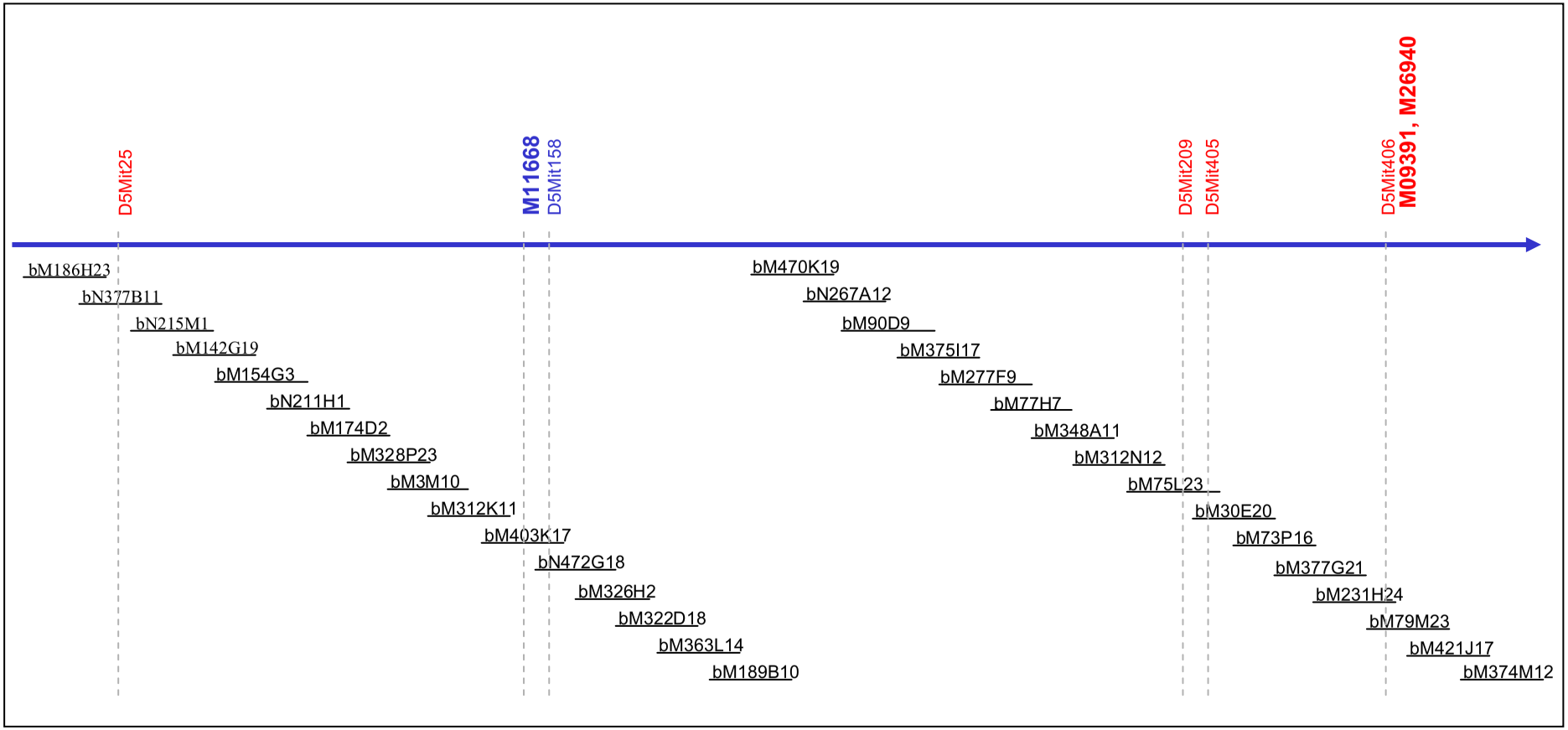



**Figure 3.13:** Physical BAC map of the *bronx waltzer* candidate region showing the positions of new markers obtained from the Mouse SNP Database in bold. Markers polymorphic for *bv* and 101 are in **red**, non-polymorphic markers are in **blue**.

#### 3.3.2.2 Sequence sampling

As a means of searching out new regions of sequence which harbour differences between the two strains used in the backcross, a number of sources of sequence were employed, both before and after the publication of the mouse genome sequence.

##### 3.3.2.2.1 BAC end sequences

The first set of sequences to be investigated for potential polymorphisms were the BAC end sequences used in the physical mapping of the region. Amplification of these by PCR using *bv* and 101 DNA as templates gave products which could be compared for size polymorphism by agarose gel electrophoresis. Of the 43 end sequences tested, only one showed a size differential between the two strains which could be detected using this method. Figure 3.14 illustrates the polymorphism identified using primers designed to the SP6 end sequence of the BAC bM277F9 (bM277F9SP6).

As an existing anchor on the physical map, bM277F9SP6 was known to lie within the critical region, situated proximal to *D5Mit209* (see Figure 3.16). Therefore the new marker was screened against the panel of mice from the *bv*/101 backcross which showed recombination between the known flanking markers. Examples of the results of this screen are shown in Figure 3.15 and a revised haplotype diagram showing its position is given in Figure 3.17.

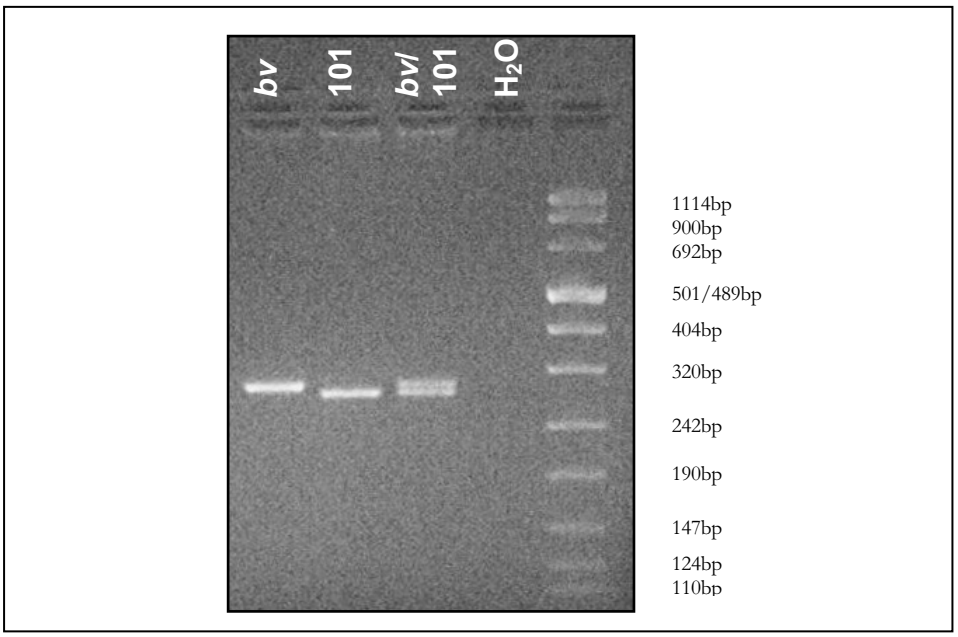



**Figure 3.14:** Agarose gel showing size polymorphism between *bv* and 101 within the sequence amplified by primers design to the SP6 end sequence of the clone bM277F9. The marker lane contains Molecular Marker VIII (Roche).

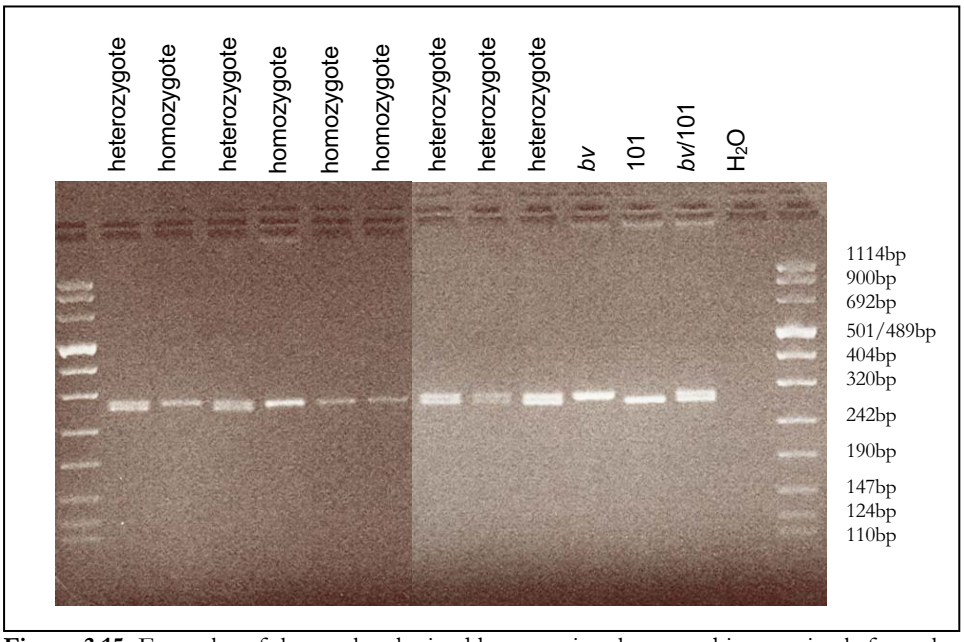



**Figure 3.15:** Examples of the results obtained by screening the recombinant animals from the *bv*/101 backcross with new polymorphic marker bM277F9SP6. Animals homozygous for the *bronx waltzer* background at this locus give a single larger-sized band. Animals heterozygous for *bv* and 101 give a double band and can be scored as having inherited 101 DNA from the F1 parent at this locus. The marker lane contains Molecular Marker VIII (Roche).

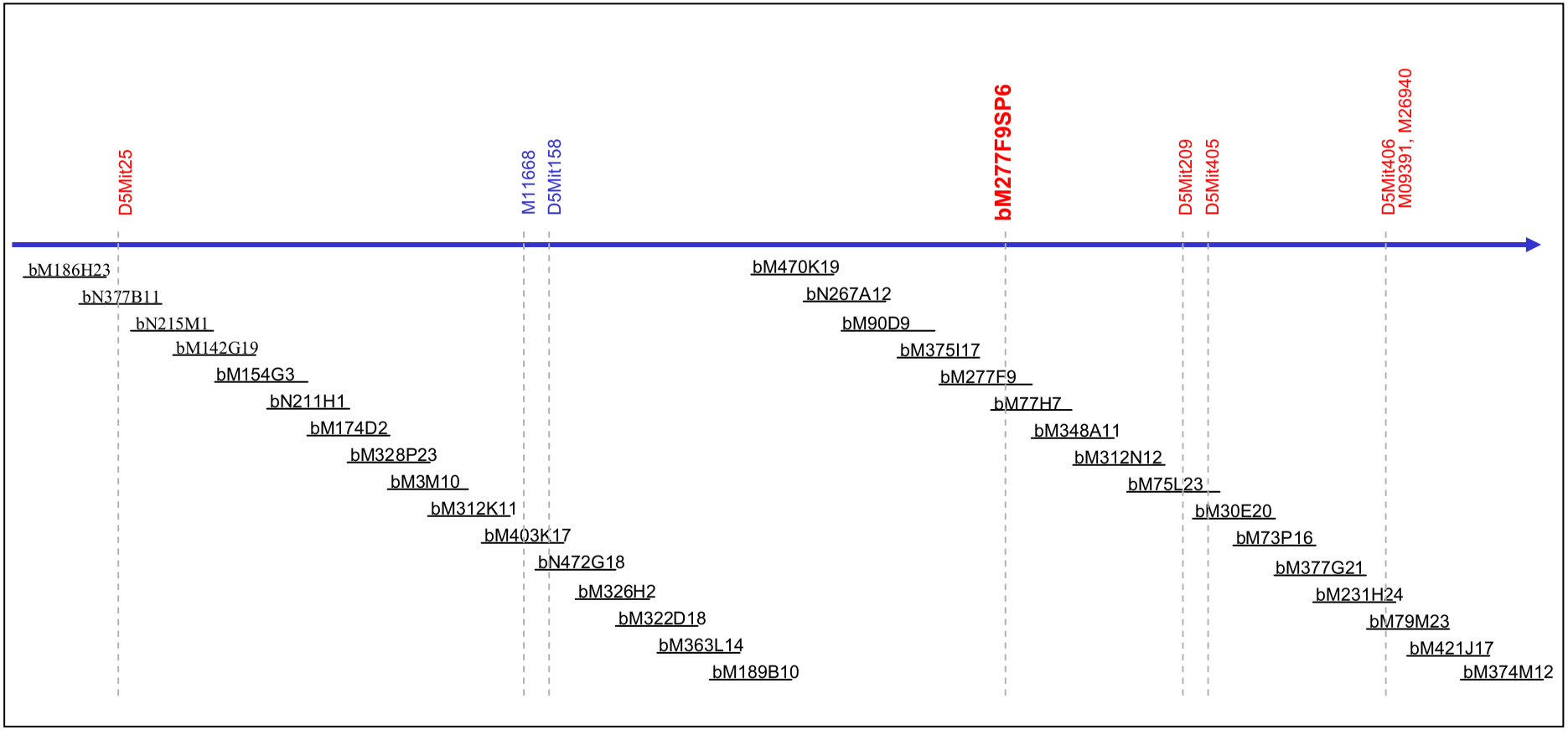



**Figure 3.16**: Physical BAC map of the *bronx waltzer* candidate region showing the position of the newly identified polymorphic marker bM277F9SP6. Markers polymorphic for *bv* and 101/H are in **red**, non-polymorphic markers are in **blue**.

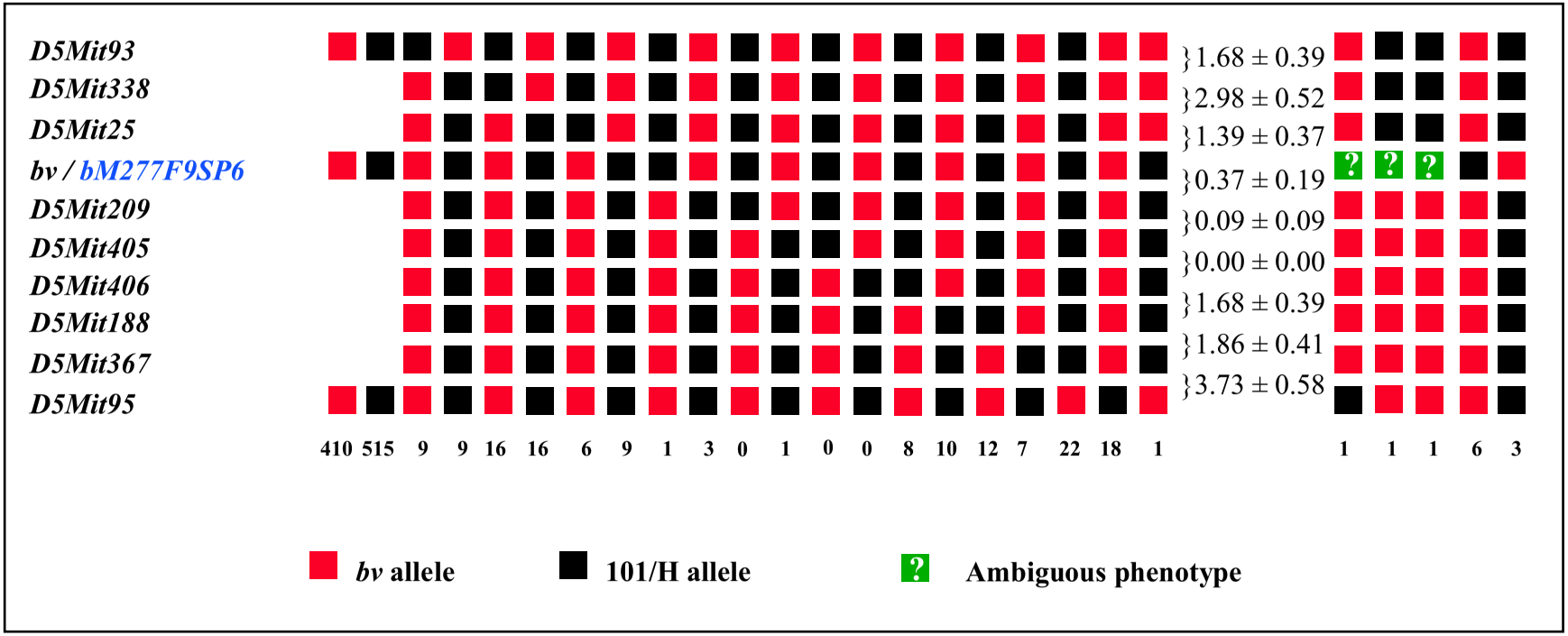



**Figure 3.17**: Haplotype analysis of the total 1085 progeny arising from the *bv*/101/H intraspecific backcross, showing the position of the newly mapped marker bM277F9SP6 (highlighted in **blue**). Genetic markers used in the mapping are listed on the left, mice typed with each pattern of recombination are shown along the bottom. Genetic distances calculated between the markers based on the number of recombinations occurring between them are given to the right. Mice excluded from the analysis as a result of discrepant or difficult phenotyping are shown on the far right of the diagram.

The new marker bM277F9SP6 was found to be non-recombinant with the *bv* locus, having in every case the same genotype as the phenotype of the backcross mouse. This result suggests that the marker lies in very close proximity to the gene responsible for the *bronx waltzer* phenotype, since for two markers to co-segregate consistently in over 1000 meioses there must have occurred no recombination between them, implying either a very short distance or a region of extremely low recombination. The recombination rate can be compared to the average rate for mouse chromosome 5, which at 150Mb (Ensembl Build 33) and 92cM (Mouse Genome Database) has an average of 1.6Mb/cM. The *bv* candidate region, at 2.8Mb and 1.86cM has an average of 1.5Mb/cM, making it very similar to the average recombination rate. Thus it is likely that bM277F9SP6 represents a very near marker for *bv*.

However, since there are no recombinations between bM277F9SP6 and *bv*, the new marker is mapped to the same place as *bv* on the genetic map (see Figure 3.18) and thus does not represent a new flanking marker. Therefore the size of the candidate region has not been reduced by the discovery of this marker, though it is possible to say that the gene is more likely to be situated in close proximity to bM277F9SP6.

One of the mice which were excluded from the original haplotype analysis, T901, appeared to show recombination between *bv* and bM277F9SP6 if the tentative phenotype is to be believed. The phenotyping notes for this mouse state that it exhibited slight head bobbing, a small Preyer reflex, reached out on landing and did not circle. At the time it was thought to most likely to be a heterozygote, meaning that it carried a 101/H copy of the *bv* allele from the F1 parent. If this were the case then the new marker bM277F9SP6 would represent a new distal marker on the genetic map, mapping just 0.09cM of the *bv* locus and reducing the size of the candidate region by 0.28cM to 1.67cM. However, should the mouse in fact carried a mutant *bv* copy of the allele then bM277F9SP6 would be mapped proximal of the *bv* locus and since the phenotype could not definitively be recorded, the mouse must be discounted from the results.

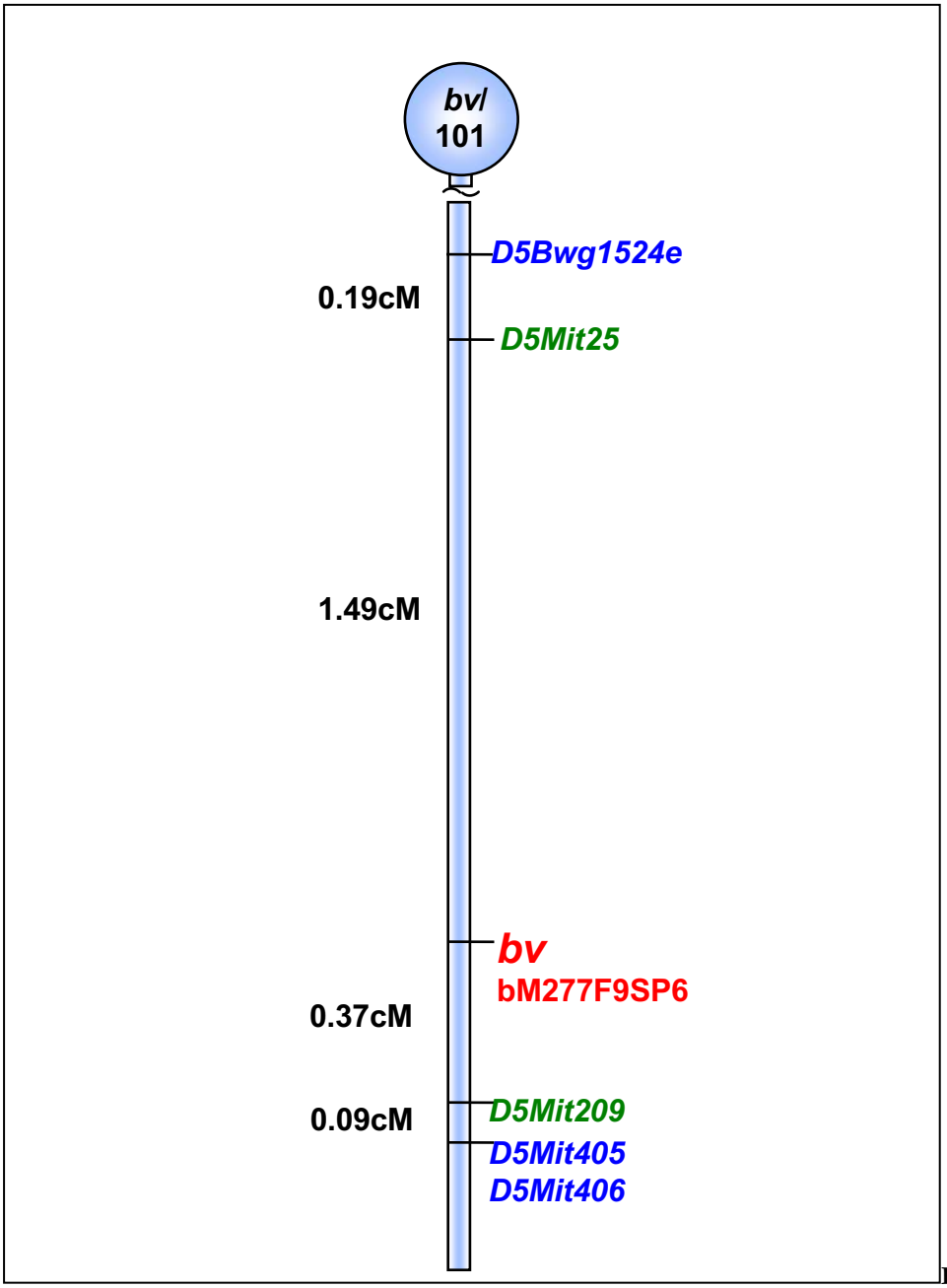



**Figure 3.12:** Genetic map of the *bv*/101 backcross showing the position of the newly mapped marker bM277F9SP6 which co-segregates with *bronx waltzer*. Genetic loci are drawn to the right of the diagram, with the *bronx waltzer* locus shown in red, flanking markers in green and other genetic markers in blue. Genetic distances determined by the number of recombinations observed between markers are given in black to the left of the diagram.

#### 3.3.2.2.2 3' untranslated regions

The second source of sequences sampled prior to the release of the sequence of the mouse genome were untranslated regions of genes annotated within the critical interval. Once again, these were amplified by PCR using *bv* and 101 templates and the sizes of the products compared by agarose gel electrophoresis. Twenty such sequences originating from 14 different genes were compared in this way, of which one proved to be polymorphic. The visible size differential between fragments amplified using primers designed to part of the untranslated region of the gene *Uracil-DNA Glycosylase* (*UNG*) is illustrated in Figure 3.19.

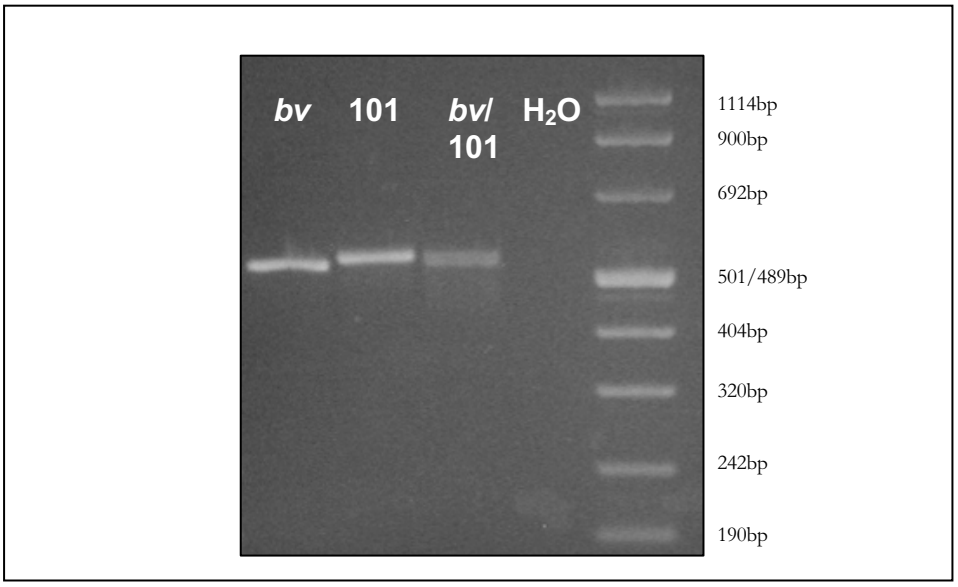



**Figure 3.19:** Agarose gel showing different sized bands resulting from amplification of the marker UNG-utr1 using *bronx waltzer* and 101/H template DNA. The marker lane contains Molecular Marker VIII (Roche).

This new polymorphic marker was used to screen the panel of recombinant mice from the *bv*/101 backcross in order to determine its position relative to the existing genetic markers (Figure 3.20). A single mouse was found to have a recombination breakpoint between the existing proximal marker *D5Mit25* and UNG-utr1, as shown in the haplotype diagram in Figure 3.21. Using this data the marker was plotted on the genetic map (Figure 3.22), as well as being tested against the panel of BAC insert

DNA so that its physical position within the region could be established. (Figures 3.23 and 3.24).

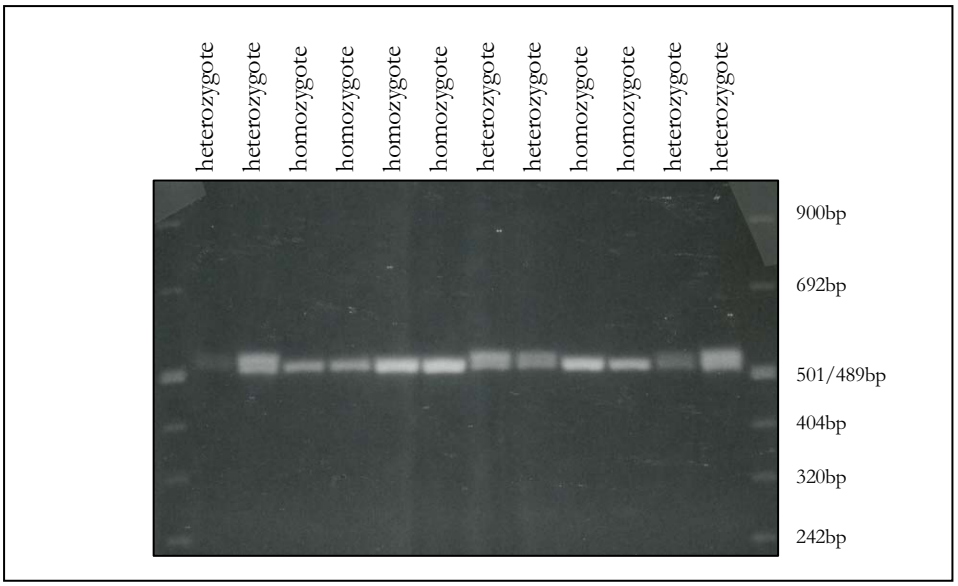



**Figure 3.20:** Agarose gels showing examples of bands amplified from backcross panel DNA using the marker UNG-utr1 to distinguish between mice homozygous or heterozygous at the locus. The marker lane contains Molecular Marker VIII (Roche).

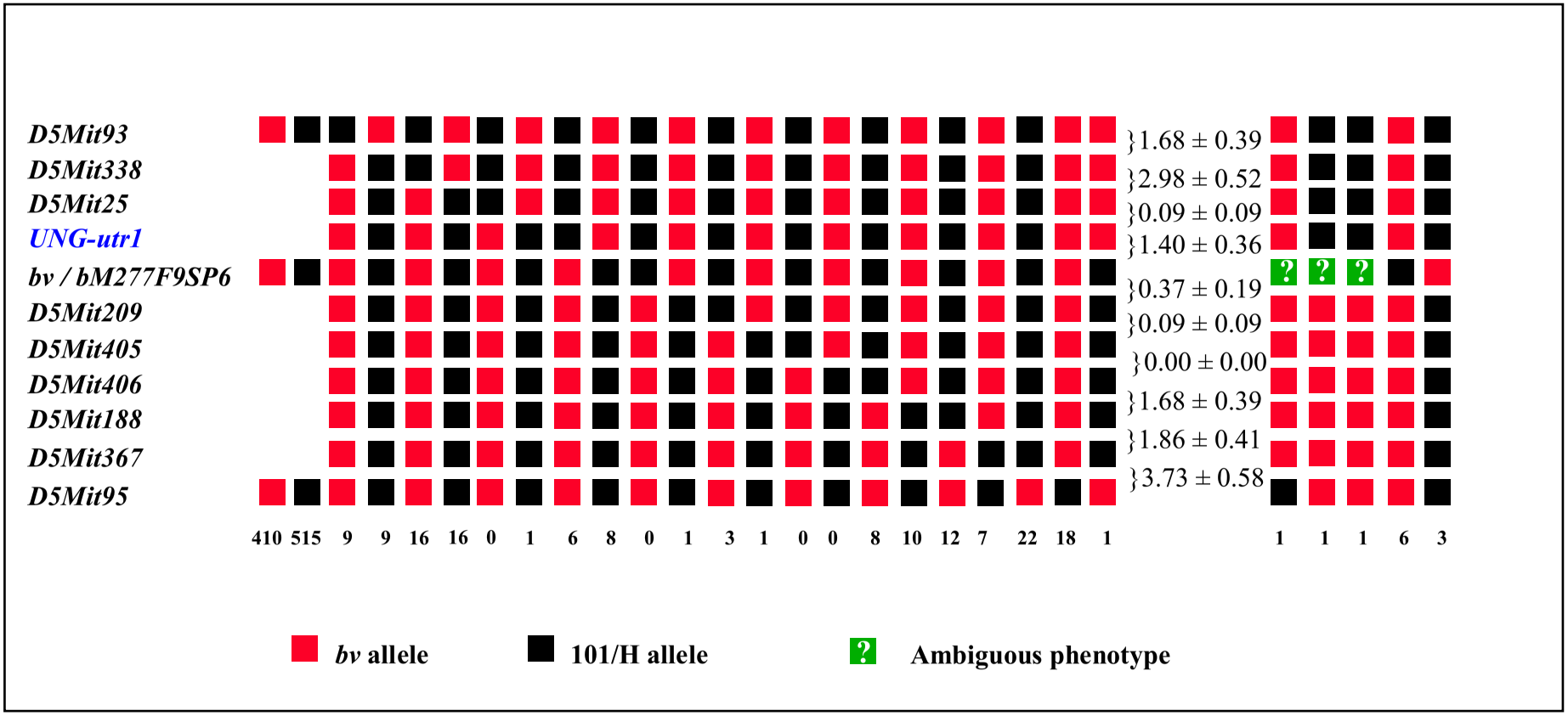



**Figure 3.21:** Haplotype analysis of the total 1085 progeny arising from the *bv*/101/H intraspecific backcross, showing the position of the newly mapped marker UNG-utr1 (in blue). Genetic markers used in the mapping are listed on the left, the number of mice typed with each pattern of recombination are shown along the bottom. Genetic distances calculated between the markers based on the number of recombinations occurring between them are given to the right. Mice excluded from the analysis as a result of discrepant or difficult phenotyping are shown on the far right of the diagram.

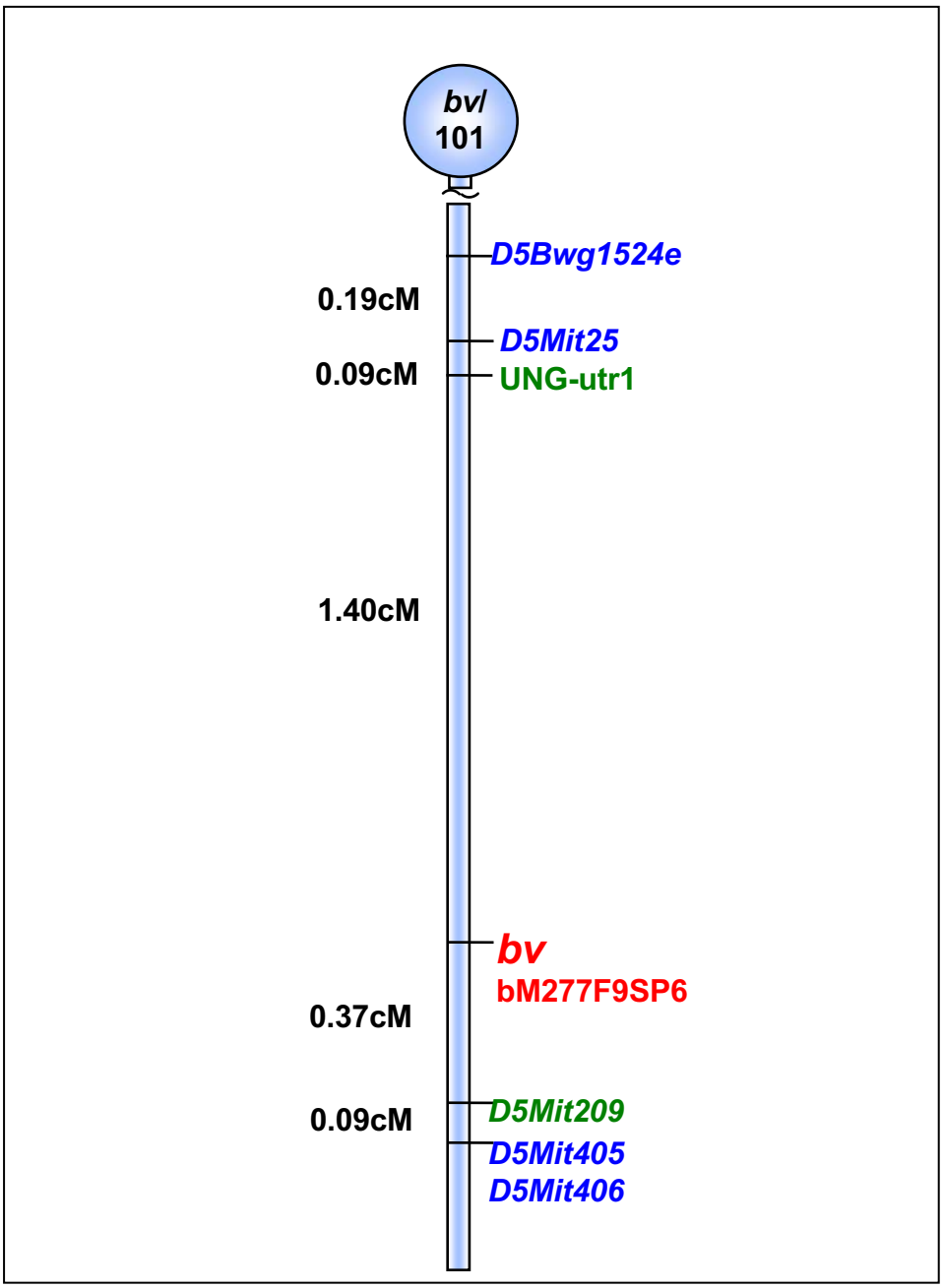



**Figure 3.22:** Genetic map of the *bv*/101 backcross showing the position of the newly mapped marker UNG-utr1. Genetic loci are drawn to the right of the diagram, with the *bronx waltzer* locus shown in **red**, flanking markers in **green** and other genetic markers in **blue**. Genetic distances determined by the number of recombinations observed between markers are given in black to the left of the diagram.

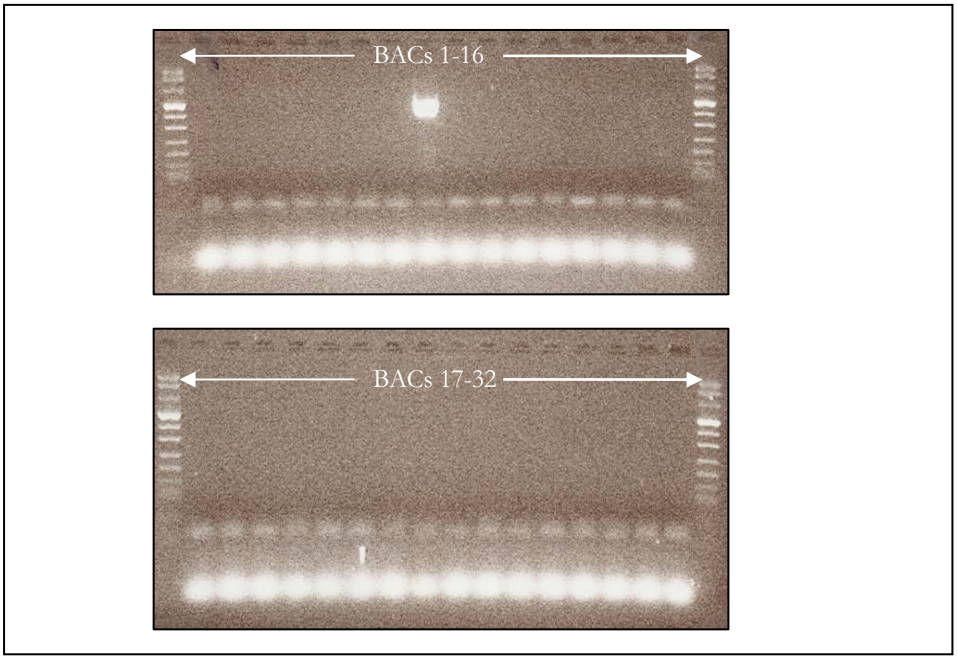



**Figure 3.23:** Agarose gels showing BAC inserts screened for the marker UNG-utr1. The positive clone is bM142G19, which is situated distal of the flanking marker D5Mit25 (see Figure 3.24).The marker lanes contain Molecular Marker VIII (Roche).

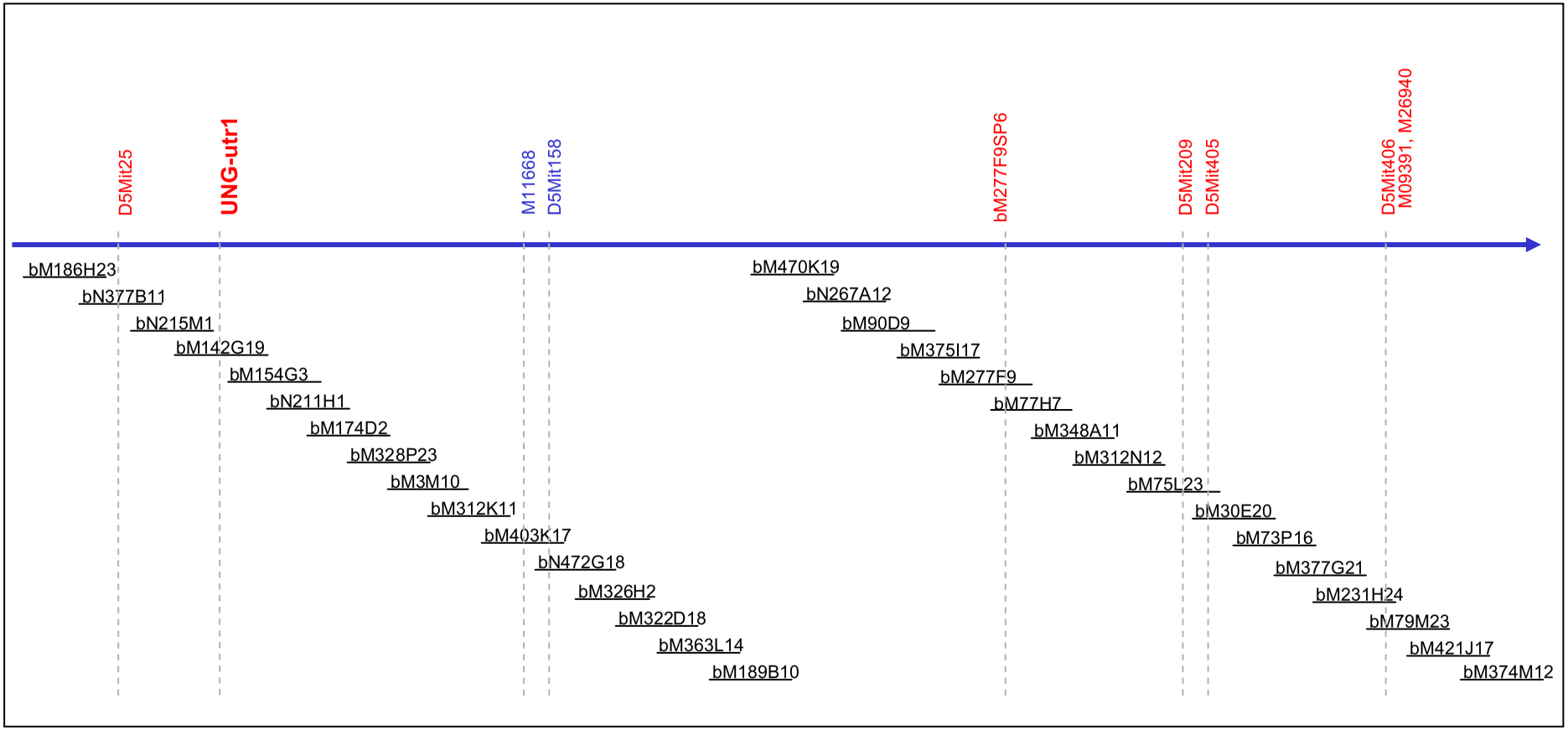



**Figure 3.24:** Physical BAC map of the bronx waltzer candidate region showing the position of the newly identified polymorphic marker UNG-utr1. Markers polymorphic for *bv* and 101 are in **red**, non-polymorphic markers are in **blue**.

UNG-utr1 was found to map between *D5Mit25* and the *bv* locus, thus establishing it as a new flanking marker for the *bronx waltzer* candidate region. Its discovery resulted in a reduction of the size of the genetic interval by 0.09cM to 1.77cM. When placed on the Mouse Ensembl physical map as shown in Figure 3.25a, it was found to reduce the physical interval by 257Kb to 2.5Mb. The establishment of this new proximal flanking marker resulted in the exclusion of seven genes from the candidate region as illustrated in Figure 3.25b. Since UNG is located on the forward strand and this marker is situated in its 3' untranslated region, the coding region and any upstream transcriptional control elements lie outside of the region defined by the new marker, and thus it too is unlikely to be a candidate gene for *bronx waltzer*.

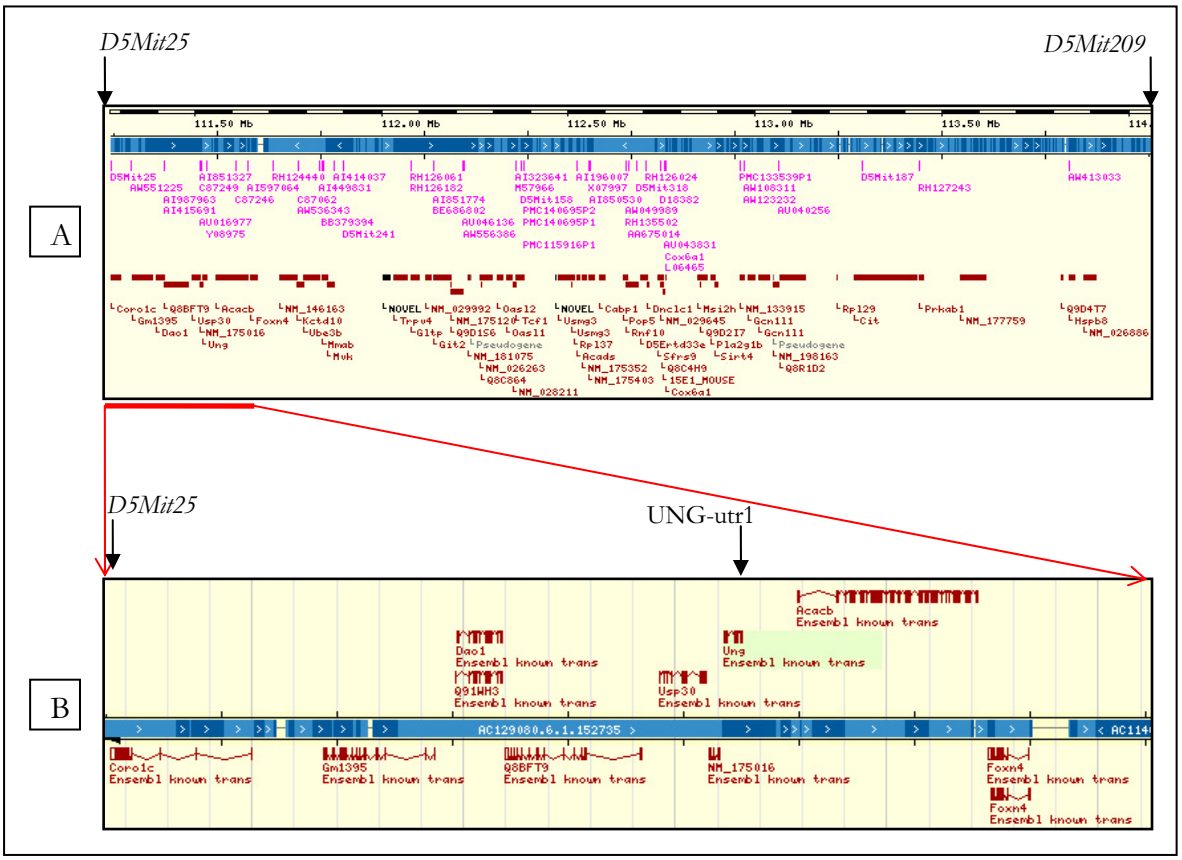



**Figure 3.25:** Screenshots from Mouse Ensembl showing the location of the newly characterised proximal flanking marker UNG-utr1. [A] depicts the original candidate region, spanning 2.8Mb from *D5Mit25* to *D5Mit209*. [B] shows the proximal end of the region in greater detail, including the placement of UNG-utr1 which reduces the size of the region by 257Kb and excludes seven genes from candidacy.

#### 3.3.2.3 Tandem repeat sequences

With the publication of the draft sequence of the mouse genome (Gregory *et al.* 2002), new sources of potential polymorphisms became available. The first of these to be exploited was the ability to locate new tandem repeats within the region, since these often vary in length between different mouse strains (Hamada *et al.* 1982). The Tandem Repeat Finder software (Benson 1999) was used to detect such sequences in the region between the flanking markers UNG-utr1 and *D5Mit209*. The data produced consisted of a repeat table containing information about each repeat, including its location, size, number of copies and nucleotide content. An example of the output is given in Figure 3.26. Using this information, repeat regions with a low period size, high copy number and high fidelity between adjacent copies were identified as being most likely to prove polymorphic since they are most likely to undergo errors in replication. A selection of 18 such repeats were chosen from areas distributed throughout the region and primers were designed to amplify these repeats. Amplification by PCR with *bv* and 101 DNA templates allowed comparison of the product sizes obtained. Examples of the results obtained in this screen are shown in Figure 3.27.

Since no obvious size differences were observed between the fragments amplified from different DNA samples, sequencing of the products was carried out in order to detect any small differences but none were identified.

| Indices | Period Size | Copy Number | Consensus Size | Percent Matches | Percent Indels | Score | A | C | G | T | Entropy (0-2) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 207577--207629 | 3 | 17.7 | 3 | 100 | 0 | 106 | 0 | 32 | 33 | 33 | 1.58 |
| 208116--208185 | 2 | 35.0 | 2 | 82 | 0 | 77 | 44 | 1 | 54 | 0 | 1.09 |
| 210862--210906 | 14 | 3.2 | 14 | 83 | 0 | 54 | 15 | 2 | 33 | 48 | 1.57 |
| 210942--211033 | 10 | 9.4 | 10 | 73 | 4 | 53 | 9 | 4 | 36 | 48 | 1.56 |
| 212461--212514 | 4 | 12.5 | 4 | 98 | 2 | 72 | 0 | 0 | 44 | 55 | 0.99 |
| 212509--212538 | 7 | 4.3 | 7 | 100 | 0 | 60 | 0 | 0 | 26 | 73 | 0.84 |
| 213458--213483 | 1 | 26.0 | 1 | 100 | 0 | 52 | 0 | 0 | 0 | 100 | 0.00 |
| 214190--214223 | 11 | 3.2 | 11 | 83 | 8 | 52 | 55 | 23 | 0 | 20 | 1.43 |
| 214276--214449 | 8 | 21.4 | 8 | 70 | 14 | 88 | 51 | 19 | 2 | 26 | 1.58 |
| 215964--216007 | 2 | 22.0 | 2 | 100 | 0 | 88 | 50 | 0 | 0 | 50 | 1.00 |
| 221664--221713 | 2 | 25.0 | 2 | 100 | 0 | 100 | 50 | 0 | 0 | 50 | 1.00 |
| 224104--224170 | 12 | 5.6 | 12 | 80 | 0 | 62 | 29 | 58 | 0 | 11 | 1.34 |
| 227640--227723 | 4 | 21.0 | 4 | 95 | 0 | 114 | 67 | 0 | 32 | 0 | 0.91 |
| 230326--230371 | 8 | 6.0 | 8 | 85 | 10 | 51 | 50 | 15 | 34 | 0 | 1.44 |
| 230341--230373 | 4 | 8.3 | 4 | 93 | 0 | 57 | 51 | 21 | 27 | 0 | 1.48 |
| 230668--230700 | 2 | 16.5 | 2 | 100 | 0 | 66 | 51 | 48 | 0 | 0 | 1.00 |
| 234119--234182 | 19 | 3.4 | 19 | 85 | 8 | 94 | 7 | 32 | 32 | 26 | 1.85 |
| 235820--235981 | 21 | 7.8 | 19 | 79 | 16 | 153 | 15 | 62 | 0 | 22 | 1.32 |
| 239107--240044 | 9 | 105.1 | 9 | 69 | 22 | 357 | 9 | 49 | 19 | 21 | 1.76 |
| 239062--240064 | 4 | 232.8 | 4 | 73 | 14 | 707 | 9 | 49 | 19 | 21 | 1.77 |
| 239065--240077 | 13 | 76.6 | 13 | 77 | 13 | 827 | 9 | 49 | 19 | 21 | 1.77 |
| 241636--241662 | 1 | 27.0 | 1 | 100 | 0 | 54 | 100 | 0 | 0 | 0 | 0.00 |

**Figure 3.26:** Example of the output obtained when analysing the candidate region for potential polymorphic repeat regions using Tandem Repeat Finder software. Repeats were selected to have a period size of between 2 and 4, a copy number great than 20 and a high percentage match between adjacent repeats. Those meeting these criteria are marked by red arrows.

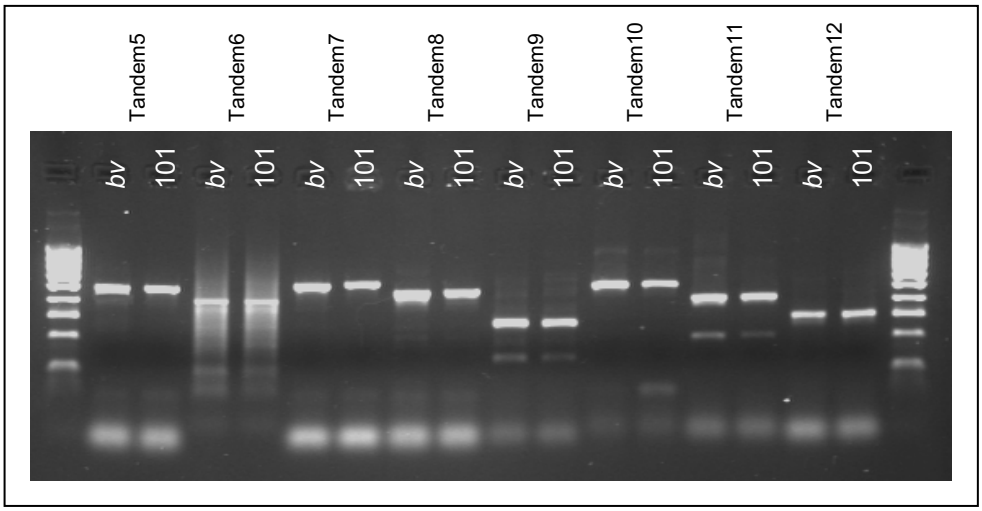



**Figure 3.27:** Examples of the sizes of products amplified from *bv* and 101 DNA using primers designed around tandem repeat sequences. Although Tandem 7 appears to give slightly different sized bands, no difference was identified when the products were sequenced. The marker lanes contain 1Kb ladder.

#### 3.3.2.4 SNPs from published sequence

##### 3.3.2.4.1 Annotated SNPs

SNPs reported between the published sequence from C57BL/6J and other strains were investigated as a potential source of SNPs between *bv* and 101/H. A total of 292 such SNPs were identified from Ensembl Build32 between the flanking markers of the *bronx waltzer* region and these were amplified using 70 sets of primers designed to flank them. In most cases a single set of primers could be used to amplify several annotated SNPs where they were clustered together. Of these, 68 were successfully amplified from *bv* and 101/H templates and 56 gave sequence of a suitable quality for analysis. However, none of them showed polymorphism between the strains.

##### 3.3.2.4.2 MICER SNPs

Another resource which became available subsequent to the publication of the genome sequence was a set of SNPs identified between the mouse strain C57BL/6J which was the subject of the public sequencing effort and the strain 129S5/SvEv$^{Brd}$ which was used in the creation of the Mouse Insertion and Chromosome Engineering Resource. A total of 187 of the SNPs highlighted in this way were found to lie between the flanking markers of the *bv* critical region and 119 sets of primers were used to amplify them. Of the amplicons sampled, 87 were successfully sequenced and of these one was shown to contain a SNP between the two strains. Sequence traces illustrating this SNP are shown in Figure 3.28.

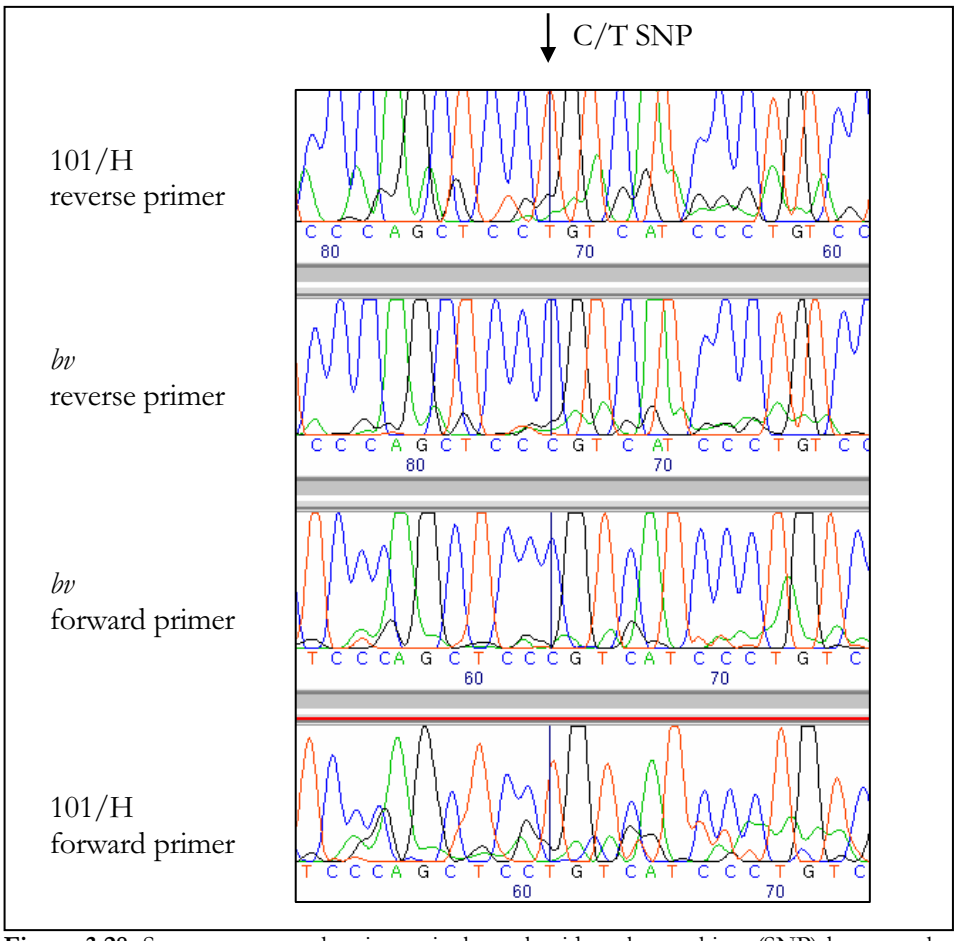



**Figure 3.28:** Sequence traces showing a single nucleotide polymorphism (SNP) between the mouse strains *bronx waltzer* and 101/H. The fragment was amplified using primers designed to a SNP identified by the MICER project, designated DASNP3.

The marker containing the SNP, DASNP3 was used to screen the backcross recombinant panel by sequencing products amplified from each mouse sample and was found to be positioned on the genetic map in the same place as UNG-utr1 (Figure 3.29).

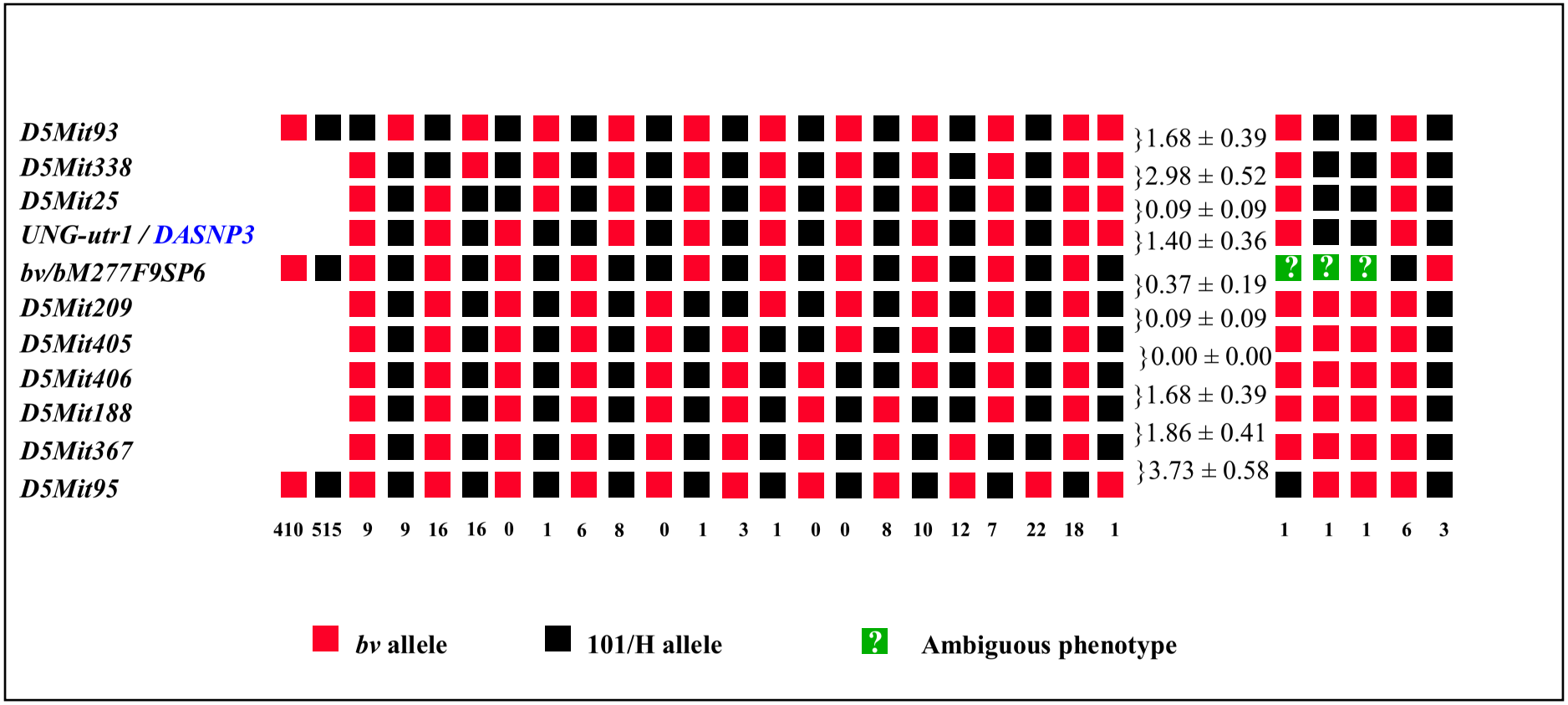



**Figure 3.29:** Haplotype analysis of the total 1085 progeny arising from the *bv*/101/H intraspecific backcross, showing the position of the newly mapped marker DASNP3 highlighted in **blue**. Genetic markers used in the mapping are listed on the left, numbers of mice typed with each pattern of recombination are shown along the bottom. Genetic distances calculated between the markers based on the number of recombinations occurring between them are given to the right. Mice excluded from the analysis as a result of discrepant or difficult phenotyping are shown on the far right of the diagram.

Although the placement of this marker does not alter the size of the candidate region on the genetic map, physically it is situated distal of UNG-utr1 (Figure 3.30), making it a new proximal flanking marker and reducing the size of the critical region by 78Kb. The marker lies within an intron of the gene *Acacb*, the most proximal gene in the region. As such, it does not fully exclude any further genes from candidacy since the mutation could still lie in an exon further downstream, but it does strengthen the case for the exclusion of *UNG*.

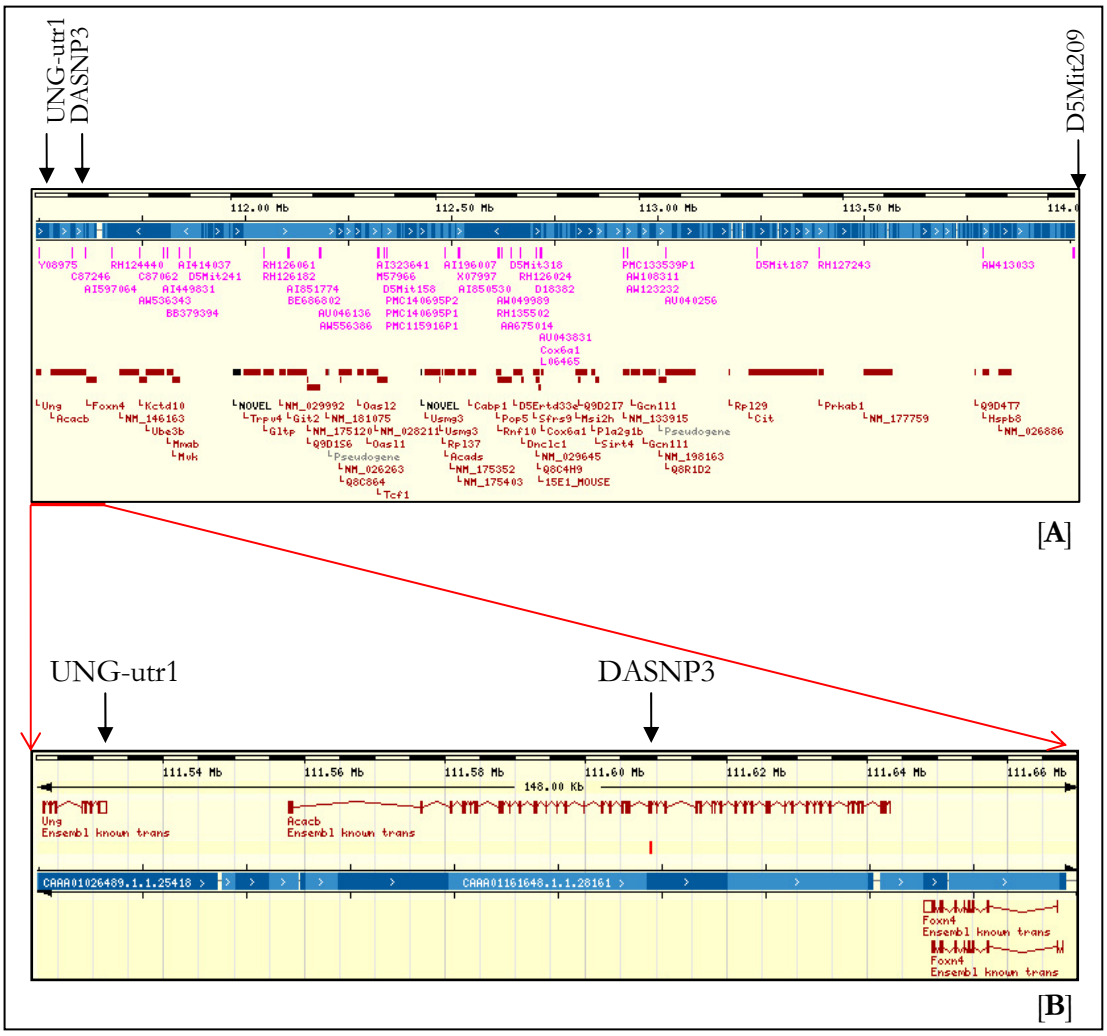



**Figure 3.30:** Screenshots from the Mouse Ensembl database illustrating the positions of markers flanking the *bv* candidate region. The entirety of the region is illustrated in [A], while [B] shows an enlarged section of the region depicting more accurately the placement of markers UNG-utr1 and DASNP3

## 3.4 DISCUSSION

Over the course of this project, much more information has become available concerning the genomic composition of the mouse. As these data were released they have been incorporated into the approaches used to further characterise and refine the candidate region for the *bronx waltzer* gene. This began with the use of physical mapping data to establish a contiguous series of BAC clones spanning the region, allowing the first estimate of its physical size. Later, newly available sequence data was employed in efforts to identify new markers polymorphic for *bronx waltzer* and 101/H. During the course of this study, a total of 422 amplicons from within the *bv* candidate region were analysed for polymorphisms between *bronx waltzer* and the backcross strain 101/H. These covered 563 potential polymorphisms, including 482 reported SNPs and 81 other sequences which carried the possibility of variation between strains. Of these only three were found to exhibit polymorphisms, of which two cosegregate and form a new proximal flanking marker reducing the size of the region by 0.09cM to 1.77cM and by 334Kb to 2.45Mb with the exclusion of seven candidate genes. The third marker is non-recombinant with the *bv* locus and therefore cannot be used to reduce the size of the region, but does carry the implication that the gene responsible for the *bronx waltzer* phenotype is likely to be situated very close by. A final version of the genetic map showing the placement of these novel markers is given in Figure 3.31, while the physical positions are illustrated in Figure 3.32.

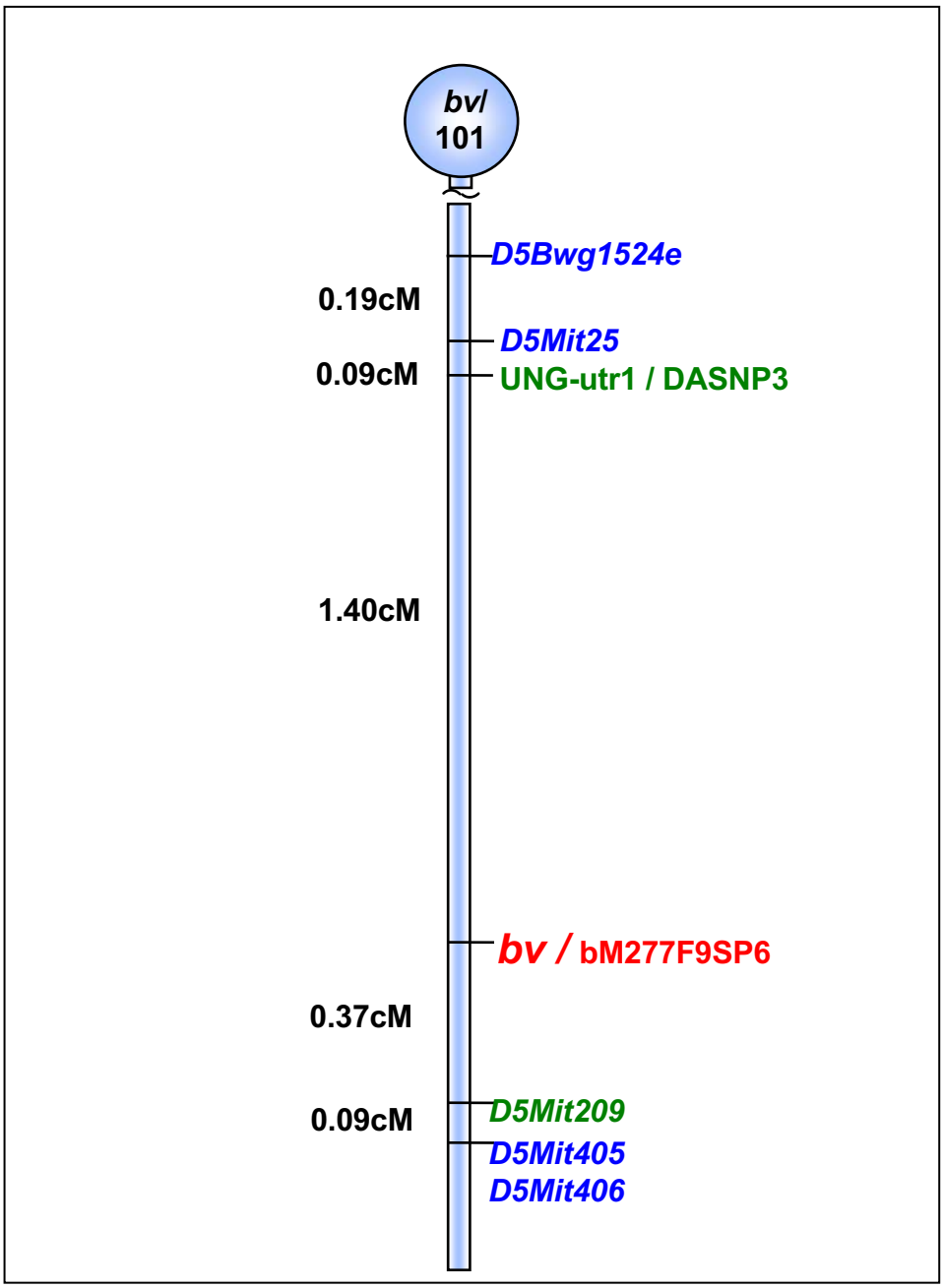



**Figure 3.31:** Genetic map of the *bv*/101 backcross showing the position of the newly mapped polymorphic markers. Genetic loci are drawn to the right of the diagram, with the *bronx waltzer* locus shown in **red**, flanking markers in **green** and other genetic markers in **blue**. Genetic distances determined by the number of recombinations observed between markers are given in black to the left of the diagram.

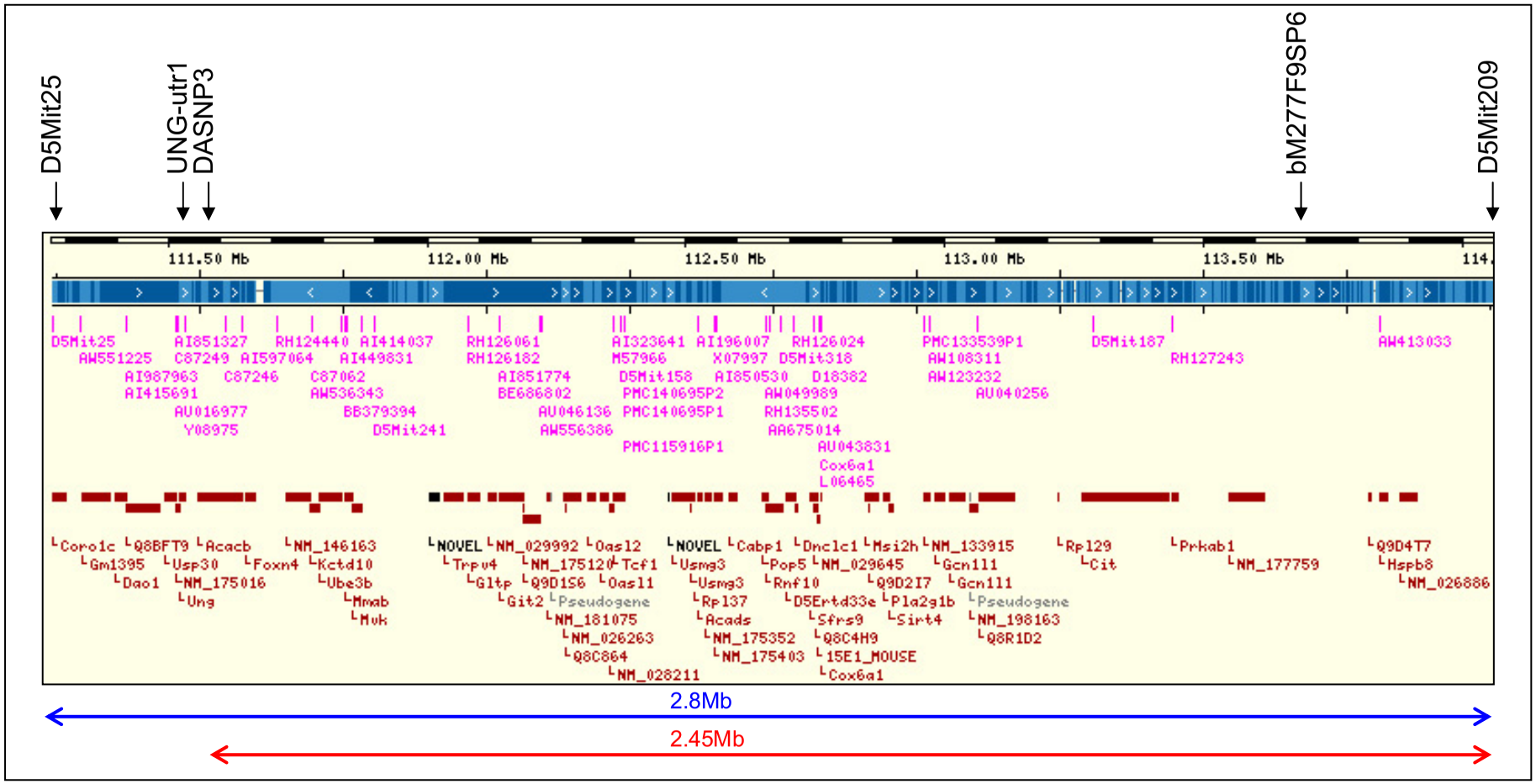



**Figure 3.32** Screenshot from Mouse Ensembl Build 33 depicting the final candidate region for the bronx waltzer locus and showing the locations of existing and newly determined markers polymorphic for *bv* and 101/H. The original candidate region was delineated by the markers D5Mit25 and D5Mit209 and is marked by the blue arrow. The new proximal marker DASNP3 reduces the size of the region by 0.35Mb and excludes seven genes from the region as shown by the red arrow. bM277F9SP6 is non-recombinant with the *bv* locus, making it unsuitable for use as a flanking marker but suggesting that the gene encoding bronx waltzer is likely to be in its vicinity.

The very low rate of polymorphisms in this region despite the high level of sequence sampling suggests that the area is conserved between the two strains being investigated, *bronx waltzer* and 101/H. Even though the inbred strain was chosen from a panel as being the most polymorphic when typed against 33 SSLP markers from mouse chromosome 5 (Bussoli *et al.* 1997), the region close to the *bv* locus will have been inherited by almost all mutant mice and so been selected for during normal breeding for maintenance of the colony, and therefore may not be representative of the rest of the chromosome. Since the background on which the *bv* mutation arose and is maintained is unknown, it is possible that this portion of its genome was inherited from a mouse strain closely related to 101/H. The degree of polymorphism between *bronx waltzer* and a number of inbred strains within this small region is investigated further in Chapter 4 as a means of choosing a more polymorphic strain for a subsequent cross, and of providing clues as to which mouse inbred strain background the *bv* mutation may have arisen on.