

## Methods

# A Sensitive Method for Detecting Variation in Copy Numbers of Duplicated Genes

Gerli Pielberg,<sup>1</sup> Andy E. Day,<sup>2</sup> Graham S. Plastow,<sup>2</sup> and Leif Andersson<sup>1,3,4</sup>

<sup>1</sup>Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden; <sup>2</sup>PIC Group, University of Cambridge, Department of Pathology, Cambridge, UK; <sup>3</sup>Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden

Gene duplications are common in the vertebrate genome, and duplicated loci often show a variation in copy number that may have important phenotypic effects. Here we describe a powerful method for quantification of duplicated copies based on pyrosequencing. A reliable quantification was obtained by amplification of the duplication breakpoint and a corresponding nonduplicated sequence in a competitive PCR assay. A comparison with an independent method for quantification based on the Invader technology revealed an excellent correlation between the two methods. The pyrosequencing-based method was evaluated by analyzing variation in copy number at the duplicated *KIT/Dominant white* locus in pigs. We were able to distinguish haplotypes at this locus by combining the duplication breakpoint test with a diagnostic test for a functionally important splice mutation in the duplicated gene. An extensive allelic variation, including the presence of a new allele carrying a single *KIT* copy expected to encode a truncated *KIT* receptor, was revealed when analyzing white pigs from commercial lines.

Duplicated genes and sequences are common in the mammalian genome. In humans, an estimated 5% of the genome consists of interspersed duplications that have arisen over the past 40 million years of evolution (Eichler 2001; Bailey et al. 2002). Many of these duplications exhibit a high degree of sequence identity at the nucleotide level (>95%), span large genomic distances, and can include both exonic and intronic sequences. In most cases duplicated blocks are composed of smaller modules, and their organization and distribution have made mapping and sequencing of these regions problematic. Sequence duplications can affect gene dosage, generation of protein diversity, and may lead to the evolution of new functions. They also increase the probability of secondary genomic rearrangements, as large blocks of identical or near identical sequence can provide substrates for subsequent nonhomologous recombination events. There are many human disorders associated with segmental duplications, and it has been estimated that 1 in every 1000 human births has a duplication-mediated germline rearrangement (Eichler 2001). Many duplicates appear to be nonfunctional, since they might lack the essential upstream regulatory sequences or a complete complement of exons. However, altered expression of duplicated genes may cause genetic disorders (Lynch 2002) or lead to functionally important polymorphism, as documented for human cytochromes (Løvlie et al. 1996; Lundqvist et al. 1999; Xu et al. 2002) and chemokines (Townson et al. 2002), and as described below for the porcine *Dominant white/KIT* locus. Genetic studies of duplicated genes are complicated because the copy number often varies between haplotypes, and they are expected to show concerted evolution due to unequal crossing-over and gene conversion (Ohta 1990).

Here we present a sensitive method for analyzing variation in gene copy number, and its application to genetic analysis of the duplicated *Dominant white/KIT* locus in pigs. *KIT* encodes the mast/stem cell growth factor receptor, which has a key role in melanocyte migration and survival (Bernstein et al. 1990; Jackson 1994). The Dominant white phenotype in pigs is caused by

two *KIT* mutations: one gene duplication associated with a partially dominant phenotype, and a splice mutation (in one of the copies) leading to the fully dominant allele (Johansson-Moller et al. 1996; Marklund et al. 1998). The splice mutation is a G to A substitution in the first nucleotide of intron 17, leading to skipping of exon 17, which is composed of 123 bp. Six *Dominant white/KIT* alleles have been identified so far: the recessive *i* allele for normal color, the semidominant *I<sup>p</sup>* allele for the Patch phenotype, the fully dominant *I<sup>1</sup>*, *I<sup>2</sup>*, and *I<sup>3</sup>* alleles for the Dominant white phenotype, and *I<sup>bc</sup>* for the dominant Belt phenotype. Alleles *i* and *I<sup>bc</sup>* carry single *KIT* copies, *I<sup>1</sup>* and *I<sup>p</sup>* carry a *KIT* duplication, and *I<sup>2</sup>* and *I<sup>3</sup>* are associated with a triplication (Johansson Moller et al. 1996; Giuffra et al. 1999; Pielberg et al. 2002). The size of the duplication is about 450 kb, and it arose by unequal homologous recombination between LINE elements flanking *KIT* (Giuffra et al. 2002). The same unique duplication breakpoint has been identified in animals carrying the *I* and *I<sup>p</sup>* alleles across breeds, implying that the *Dominant white* and *Patch* alleles are descendants of a single duplication event (Giuffra et al. 2002).

## RESULTS

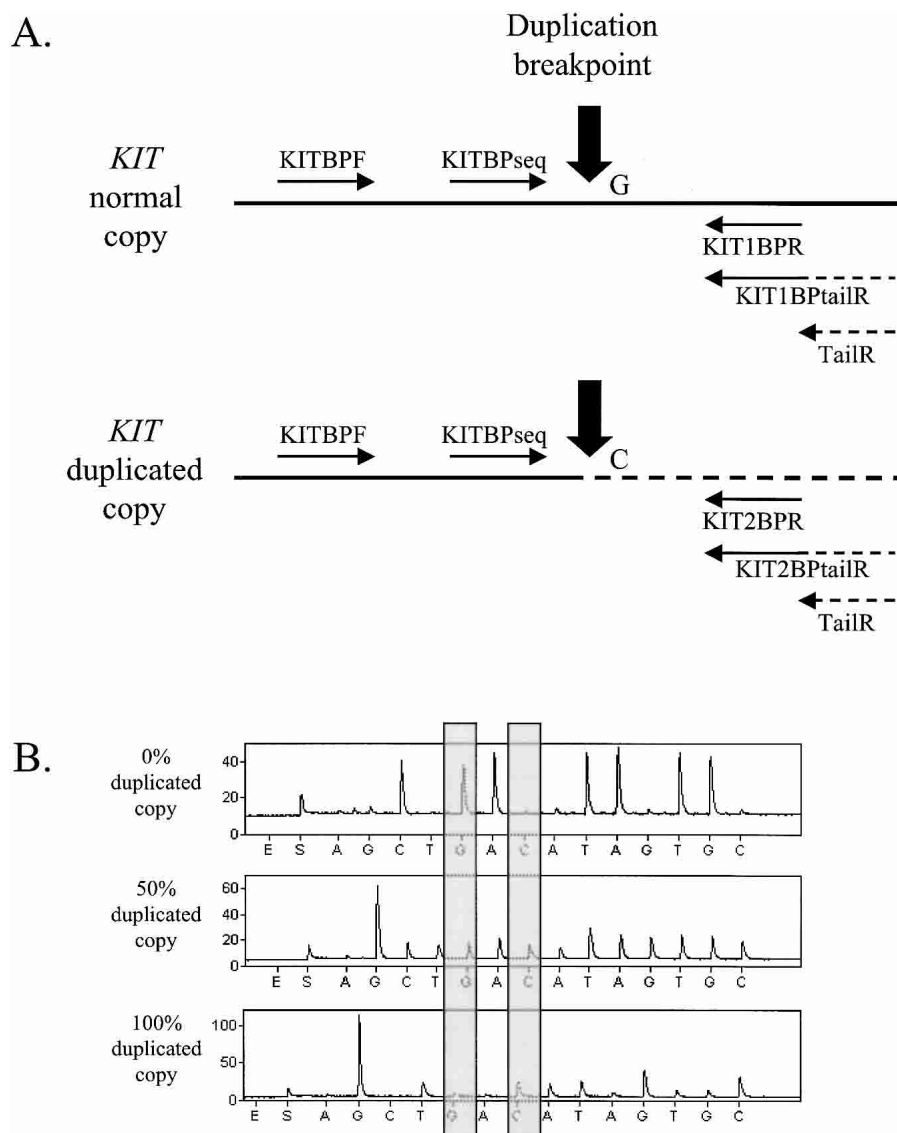
### Design of the Method for Detecting Variation in Copy Numbers of Duplicated Genes

A tandem duplication has one unique DNA sequence not present on nonduplicated chromosomes, namely the duplication breakpoint (Fig. 1A). At the breakpoint, the 3' end of the duplication is followed by the 5' end of the next copy of the duplication. We decided to take advantage of this feature and designed a competitive PCR assay to amplify fragments of 189 bp from the duplicated and normal *KIT* copy in white pigs. We used one shared forward primer (KITBPF) and two reverse primers specific for the two different *KIT* copies (KIT1BPR and KIT2BPR). Pyrosequencing using a shared sequencing primer (KITBPseq) was then used to quantify the two copies based on the relative proportion of diagnostic nucleotides after the breakpoint (Fig. 1). Pyrosequencing was chosen for this application, as previous studies have indicated that this method provides accurate quantification of different templates amplified in the same reaction (Neve et al. 2002; Pielberg et al. 2002; Wasson et al. 2002). Two different BAC

#### <sup>4</sup>Corresponding author.

E-MAIL [leif.andersson@imbim.uu.se](mailto:leif.andersson@imbim.uu.se); FAX 46 18 471-4833.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.1188203>.



**Figure 1** (A) Design of a DNA test for quantification of the gene copy number at the porcine *KIT* locus. The normal copy was PCR-amplified using the shared forward primer (KITBPF) and the specific reversed primer KIT1BPR, or alternatively a mix (1:100) of the two reversed primers KIT1BPtailR and TailR. Similarly, the duplication breakpoint was amplified using KITBPF and KIT2BPR or a mix (1:100) of KIT2BPtailR and TailR. The primer KITBPseq was used for pyrosequencing. (B) Pyrogram obtained when mixing BAC clone 1041B3 containing a nonduplicated *KIT* copy and increasing proportions of BAC clone 953F11 including the duplication breakpoint. The C and G nucleotides used for quantification are boxed.

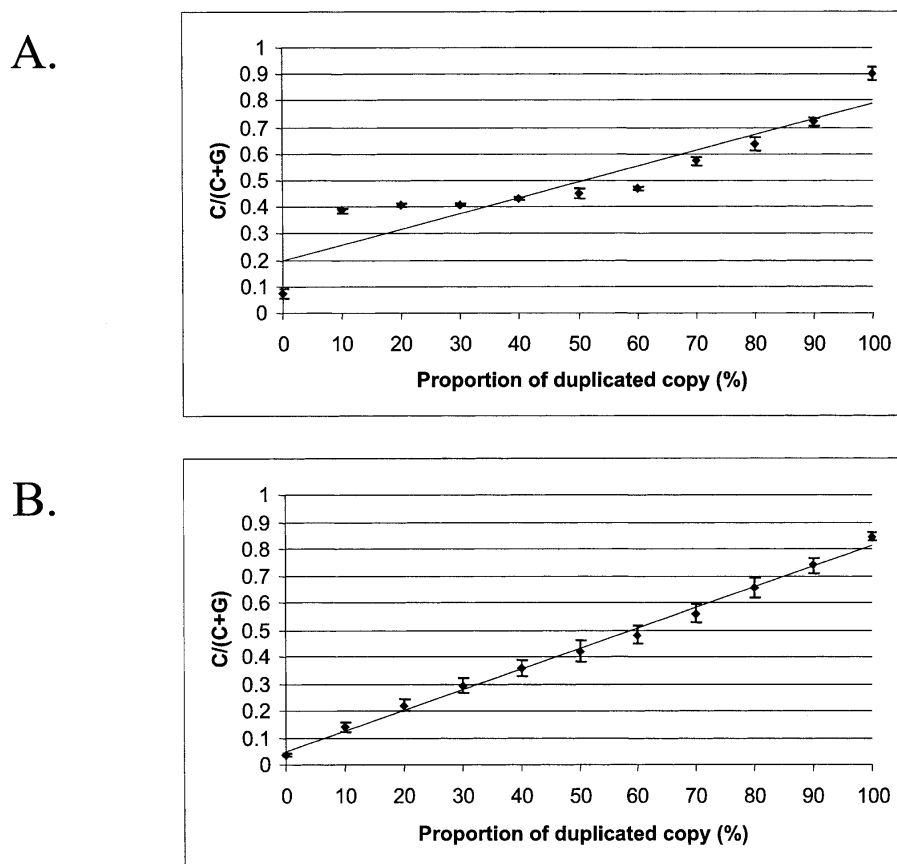
clones representing the normal copy and duplicated copy were used for calibrating the test. A standard curve was run using a dilution series containing from 0% to 100% of the duplicated copy of *KIT* (BAC 953F11) mixed with the nonduplicated copy (BAC 1041B3) of the gene.

Initially the assay was carried out with untailed primers (Fig. 2A). The resulting curve was obviously nonlinear and showed that the rarer amplicon was preferentially amplified compared to the predominant one. Because this is an end-point analysis, the nonlinearity is most likely explained by PCR kinetics. The amplification of the more common amplicon becomes retarded, as the concentration of the specific reverse primer falls more rapidly than the concentration of the reverse primer for the rarer amplicon. Our solution to this problem was to use a tailed-primer

approach. We added a specific tail-sequence to the 5'-ends of both reverse primers and included a third shared reverse primer (TailR), corresponding to the tail-sequence, in the PCR reaction (Fig. 1A). Our idea was to use the specific reverse primers in a considerably lower amount, so that their role would be limited to the initiation of PCR. The two amplicons would then compete for the same forward and reversed primers (KITBPF and TailR) during the reaction. The shared reverse primer was used in the same amount as the forward primer (5 pmol), whereas the specific reverse primers were used at concentrations of 0.05, 0.005, and 0.0005 pmol giving a 100, 1000, and 10,000 times excess of the shared primers, respectively. We observed very similar results using 0.05 and 0.005 pmoles of the specific reverse primers, whereas the reactions using 0.0005 pmoles gave too little PCR-product for successful pyrosequencing analysis. Based on these results we decided to use 0.05 pmoles of the specific reverse primers when analyzing genomic DNA. The results from the tailed-PCR approach showed a very good linearity, with the correlation coefficient being as high as 0.997, proving the accuracy of the method (Fig. 2B).

### Evaluation of the Method Using Genomic DNA and Comparison With an Invader Assay

The accuracy of our copy number test based on pyrosequencing was evaluated using a pedigree material that has been carefully studied for variation in the number of gene copies. We previously estimated the number of *KIT* copies in each individual of a wild boar/Large White intercross pedigree using Southern blot analysis, a quantitative analysis of the *KIT* splice mutation, and family segregation analysis (Johansson Moller et al. 1996; Pielberg et al. 2002). The results of the copy number test from two replicates of each sample were plotted against each other, using symbols corresponding to the predicted copy number at the *KIT* locus (Fig. 3A). The reproducibility of the test was very good (correlation coefficient between replicates = 0.994), and there was a strong correlation with the predicted copy number. There was a slight overlap between groups with three and four copies and between those with four and five *KIT* copies per genotype. However, the clear difference between the two latter groups (four vs. five copies) confirms our previous postulation of the existence of alleles with a *KIT* triplication. The fact that the results within each class of copy number tended to form a line rather than a swarm suggests that variations in DNA quality tend to influence the outcome of the test. Thus, there would probably be more to gain by using replicate DNA preparations from each individual than by running replicate PCR reactions from the same DNA preparations. It



**Figure 2** Standard curve of the copy number test using different proportions of two BACs (with or without the duplication breakpoint) as template. (A) Results obtained with untailed PCR primers; correlation coefficient = 0.925. (B) Results obtained with tailed PCR primers; correlation coefficient = 0.997.

is also possible that the variation within class partly reflects genetic differences due to sequence polymorphism within primer sites or within the amplified region in some copies in some haplotypes. This is in fact clearly indicated for the class with five *KIT* copies that appears to form two subgroups perfectly corresponding to the genotype classes  $I^2/I^2$  and  $I^1/I^2$  (Fig. 3A; see also Fig. 4B).

An additional confirmation of the accuracy of the pyrosequencing test was obtained in an analysis of 108 pigs from commercial white lines with both pyrosequencing and a biplex assay based on the Invader technology (Hall et al. 2000; Lyamichev et al. 2000; Neville et al. 2002; Olivier et al. 2002). The Invader assay is considered to be useful for quantifications, because it involves a linear amplification of the target sequence, whereas a PCR reaction involves an exponential amplification. The comparison revealed an excellent correlation between the two methods (Fig. 3B).

### The Analysis of *KIT* Genotypes in a Wild Boar/Large White Pedigree

The copy number test, based on pyrosequencing, in combination with the quantitative splice test (Pielberg et al. 2002) was applied to determine *KIT* genotypes in our wild boar/Large White intercross. The two wild boar founders were both homozygous  $i/i$  for the wild-type allele, whereas the eight Large White founders carried different genotype combinations of the  $I^{be}$ ,  $I^p$ ,  $I^l$ , and  $I^2$  alleles; the constitution of porcine *KIT* alleles is summarized in Table 1. By combining the copy number test and the splice test,

we were able to resolve most genotypes, and the results from the analysis of the wild boar pedigree were in good agreement with the theoretical outcome (Fig. 4A,B). The addition of the copy number test makes it possible to separate genotypes that were indistinguishable with the splice test only. For instance, genotypes  $i/i$ ,  $I^{be}/i$ , and  $I^p/i$ , all show the same percentage of splice versus normal, that is, 0%. We can now easily distinguish the  $I^p/i$  genotype, because it carries the duplication breakpoint that is not present in the other two genotypes.

### Extensive Allelic Diversity in Commercial White Populations

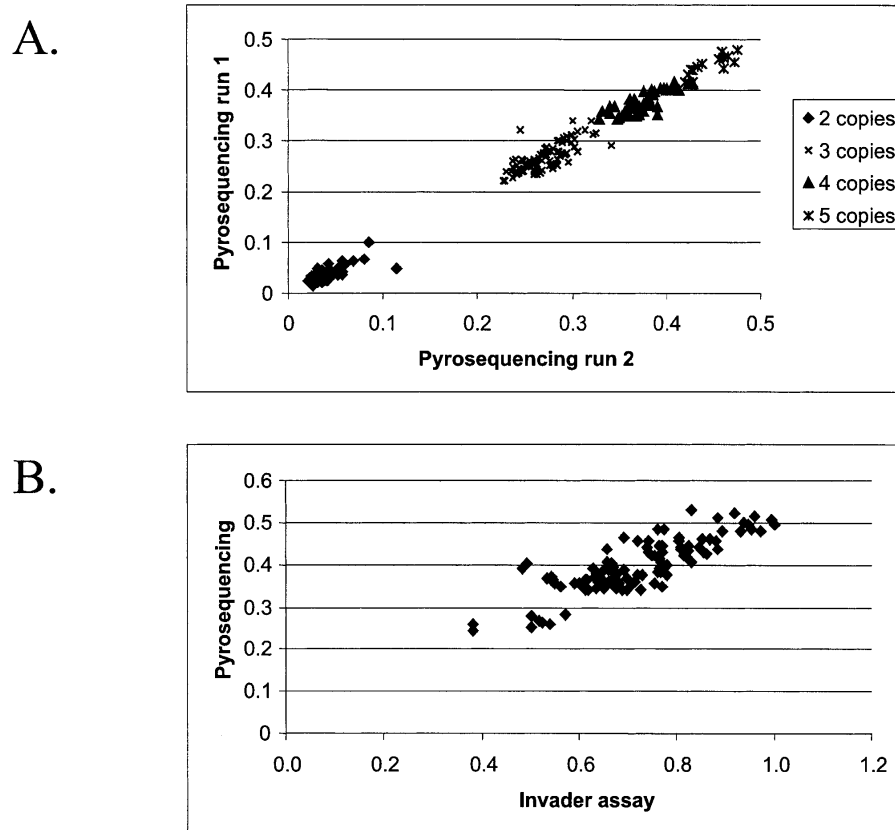
The analysis of pigs from a synthetic line based on the Large White breed, but still segregating for coat color ( $n = 187$ ) revealed an extensive allelic diversity at the *KIT* locus (Fig. 4C). Most genotype combinations of the  $i/I^{be}$ ,  $I^p$ ,  $I^l$ , and  $I^2$  alleles were identified. Interestingly, a new allele was discovered in this material. This allele has a single *KIT* copy with the splice mutation and no normal copy (Fig. 4C). The allele was denoted  $I^l$  because we assume that the allele is lethal in the homozygous condition, based on comparative data on null mutants in the mouse (Jackson 1994). The allele was found in two genotype combinations in this material  $I^{be}$  (or  $i$ )/ $I^l$  (two copies, 50% splice) and  $I^l/I^l$  (three copies, 66% splice; Fig. 4C). This allele was also detected in several commercial white populations (data not shown). We

have not yet observed animals homozygous for the  $I^l$  allele, but we are currently setting up test crosses to test the hypothesis of lethality.

The combination of the splice test and the breakpoint test allowed us to evaluate the resolution obtained with the pyrosequencing-based test compared to the Invader assay (Fig. 4D). The results indicated that the pyrosequencing test gave a slightly better resolution than the Invader assay as carried out in this study (the groups tended to be less tightly clustered in the vertical dimension with the Invader assay; cf. Fig. 4C,D). However, both methods gave sufficient resolution to assign most individuals to the correct genotype class when combined with the splice test.

### DISCUSSION

We have here described a powerful method for quantification of the number of copies of duplicated genes based on pyrosequencing. The method gave clearly more accurate results than the previously used real-time PCR amplification of *KIT* and a single-copy control sequence (Pielberg et al. 2002). The method described is simple to perform but requires prior information of the duplication breakpoint. It is particularly useful for distinguishing individuals with two copies from those with three or more copies of the duplicated region, as no signal from the breakpoint test is obtained in individuals lacking the duplication. The method also gives a good resolution to distinguish individuals with three, four, or five copies, but as the copy number increases it becomes gradually more difficult to accurately distinguish genotype



**Figure 3** Evaluation of the diagnostic test for variation in gene copy number. (A) Application of the diagnostic test for quantification of *KIT* copy numbers in a wild boar/Large White intercross (founders,  $n = 10$ ;  $F_1$ ,  $n = 22$ ;  $F_2$ ,  $n = 177$ ). The results obtained for the 1st and 2nd replicate analysis of each sample are presented. The plotted value represents the  $C/(C+G)$  ratio, i.e.,  $C$  being specific for the duplicated copy and  $G$  specific for the nonduplicated copy of the gene. The symbol used for each sample represents the copy number predicted on the basis of previous results. (B) Comparison of the estimation of *KIT* copy numbers for 108 pigs from white commercial lines using pyrosequencing technology and the Invader assay. The results given for the Invader assay are the ratio obtained by dividing the net *KIT* signal (minus background) by the net *ESR* signal (minus background).

classes that differ by only one copy. This is because the relative increase in the signal from the breakpoint becomes smaller. The Invader technology tested in this study also appears promising for this type of quantification. The Invader technology requires more genomic DNA and careful quantification of the DNA template, but does not require prior knowledge of the duplication breakpoint. In a recent study of the variation in copy number at the human *CYP2D6* locus based on the Invader technology, it was reported that the method could resolve individuals with one, two, or three copies, but it could not resolve copy numbers greater than three (Neville et al. 2002).

We also demonstrated how the allelic heterogeneity at a duplicated locus can be resolved by combining the breakpoint test with a test for a sequence polymorphism. The combined test can be utilized to assign a genotype to an unknown sample, including the likelihood for the assignment by estimating confidence intervals for each genotype class. We were able to distinguish most genotypic combinations of the seven alleles at the porcine *KIT* locus. The only exceptions included a few genotype classes that have the same gene copy number and the same splice ratio but where the haplotype composition differs between genotypes. For instance, the genotypes  $I^1/I^2$  and  $I^3/I^4$  both possess three normal and two splice copies. In these cases the genotype must be deduced by family segregation analysis. Further, we can-

not yet distinguish the  $i$  and  $I^{Be}$  alleles, both having a single *KIT* copy and no splice mutation, because the causative mutation for  $I^{Be}$  has not yet been identified.

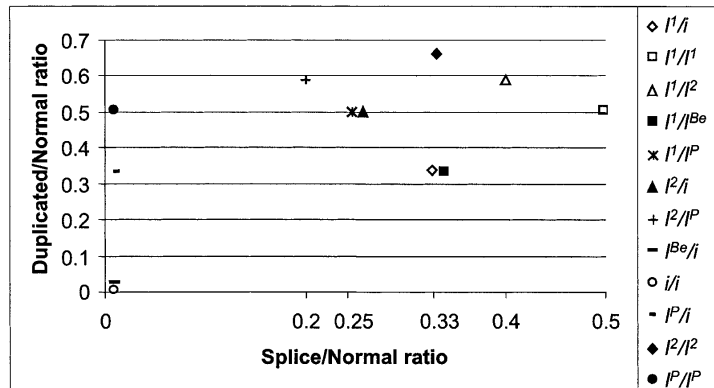
Our screening of pigs from commercial white lines has revealed an extensive allelic diversity (this study; Pielberg et al. 2002), despite the strong selection for white color for more than 100 years. We found four fairly common alleles in the lines we tested. These are  $I^1$  and  $I^2$  with the duplication and the splice mutation and associated with Dominant white color,  $I^p$  with the duplication but not the splice mutation and associated with the Patch phenotype, and a fourth allele without the duplication and without the splice mutation. This fourth allele may be  $I^{Be}$  for the Belt phenotype, the wild-type allele ( $i$ ), or another not yet characterized allele (Pielberg et al. 2002). The most likely explanation for the high allelic diversity is that unequal crossing-over, and possibly gene conversion, occur quite frequently, as expected for nearly identical tandem duplications (Ohta 1990). Balancing selection may also influence the allele frequency distribution at this locus. The splice mutation is associated with the fully dominant alleles that are favored by strong selection for white color. Despite this, the frequency of the  $I^1/I^1$  homozygotes has been less than 25% in the white lines we investigated. The *KIT* copy carrying the splice mutation is probably encoding a kinase-dead receptor, because the splice mutation leads to skipping of exon 17 that encodes 41 amino acids of an essential part of the kinase domain. Thus, the splice copy

may be potentially associated with negative pleiotropic effects on hematopoiesis and fertility, as is the case for loss-of-function mutations at the mouse *Kit/Dominant white spotting* locus (Jackson 1994). In fact, we reported earlier that  $I^1/I^1$  homozygotes in the  $F_2$  generation of our wild boar intercross were associated with a lower number of white blood cells compared to  $I/i$  and  $i/i$  animals (Marklund et al. 1998). Therefore, it is possible that the  $I^p$  and  $I^{Be}$  (or  $i$ ) alleles, both lacking the splice mutation, are maintained at a fairly high frequency despite unfavorable effects on pigmentation. Similarly, the rare  $I^3$  and  $I^4$  alleles may be generated by unequal crossing-over but kept at a low frequency by purifying selection, since they have a too-high proportion of *KIT* copies with the splice mutation (66% and 100%, respectively); the  $I^4$  allele is expected to be lethal in the homozygous condition, because it most likely lacks *KIT* signaling.

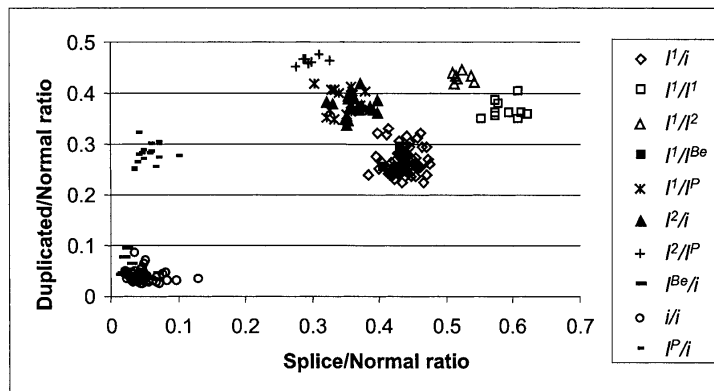
The development of a sensitive assay for analyzing haplotype diversity at the porcine *KIT* locus opens up an excellent opportunity to study the dynamic evolution of this locus. We plan to screen a large number of progeny (1000 or more) from matings between white ( $I/I$ ) and colored ( $i/i$ ) parents for mutational events changing the copy number or the ratio of the splice mutation. Because this type of mating is very common in commercial pig production, it is possible to collect the sample size needed to detect mutations. This will allow us to estimate the

## A DNA Test for Variation in Gene Copy Numbers

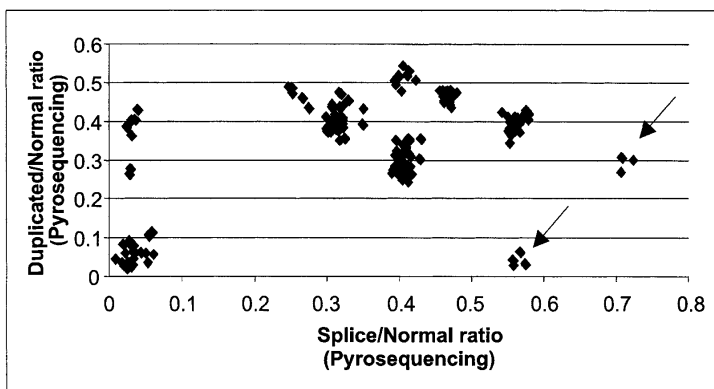
A.



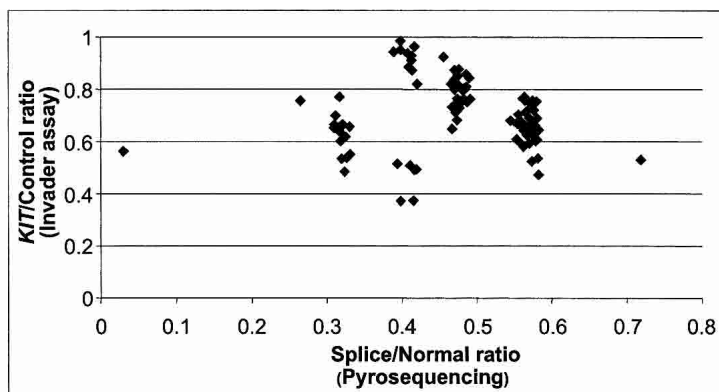
B.



C.



D.



**Figure 4** The combination of two different tests designed for the analysis of porcine *KIT* alleles. (A) The theoretical outcome. The genotypes included in the figure are all possible combinations of *KIT* alleles. The x-axis presents the ratio of splice mutation vs. normal, and the y-axis presents the ratio of duplicated vs. nonduplicated copies. Different genotypes are indicated by different symbols. (B) The analysis of pigs from a wild boar/Large White intercross: founders,  $n = 10$ ;  $F_1$ ,  $n = 22$ ;  $F_2$ ,  $n = 177$ . The x-axis shows the ratio  $[A/(A+G)]$  of splice mutation (A nt) to normal (G nt) at the first nt of intron 17 in the *KIT* gene. The y-axis shows the ratio  $[C/(C+G)]$  of duplicated copy (C nt) to nonduplicated copy (G nt) of the gene. (C) Analysis of 187 pigs from a synthetic white line (originally established by crossing Large White and Meishan breeds) segregating for color and suggesting the presence of most genotype classes predicted in (A). In addition, there is evidence of the existence of a new allele designated  $I^l$ , because it has one *KIT* copy only but with the splice mutation. There were two groups of individuals presumably heterozygous for  $I^l$  (indicated with arrows); one group with two copies and 50% splice ( $i$  or  $I^{Be}/I^l$ ) and another group with three copies and 66% splice ( $I^1/I^l$ ). (D) Combined analysis of the splice test and the Invader-based assay for estimating the number of *KIT* copies. The samples ( $n = 108$ ) were from several commercial white lines.

**Table 1.** Description of Alleles at the *Dominant white/KIT* Locus in Pigs

Allele	Total no. of KIT copies	No. of normal copies	No. of splice copies	% splice
<i>i</i>	1	1	0	0
<i>i</i> <sup>Be</sup>	1	1	0	0
<i>i</i> <sup>P</sup>	2	2	0	0
<i>i</i> <sup>1</sup>	2	1	1	50
<i>i</i> <sup>2</sup>	3	2	1	33
<i>i</i> <sup>3</sup>	3	1	2	66
<i>i</i> <sup>4</sup>	1	0	1	100

frequency of unequal crossing-over as well as gene conversion in this system.

The method described in this study has wide applications in other species, as duplicated genes are common in many species and variation in copy numbers may have important phenotypic effects. For instance, it is well documented that the human *CYP2D6* locus shows a considerable variation in copy number that is associated with the rate of metabolism of many therapeutic drugs and xenobiotics (Løvlie et al. 1996). About 7% of Caucasians are fast metabolizers because they have more copies of the gene. Genotyping of *CYP2D6* is thus of clinical relevance for adjusting the suitable dose of certain drugs. There are a number of other loci associated with human disorders for which there is a clear interest to establish a sensitive assay for studying variation at duplicated loci. These include, among many others, *CYP21A2*, associated with congenital adrenal hypoplasia (Koppens et al. 2002) and *OPN1MW*, associated with deutan colorblindness (Nathans 1999).

## METHODS

### Animals

An intercross pedigree comprising two European Wild Boar and eight Large White founders, 22 F<sub>1</sub>, and 200 F<sub>2</sub> animals was used. This pedigree has been extensively used for studies on coat color genetics (e.g., Johansson Møller et al. 1996; Kijas et al. 1998, 2001; Marklund et al. 1998; Pielberg et al. 2002). Samples were also collected from commercial white lines consisting of approximately equal numbers from seven different white dam lines: three Large White lines, two Landrace lines, and two synthetic populations based on Duroc crossed to white lines. The latter two have been selected over several generations to be fixed for dominant white. Samples were also collected from a synthetic line established several generations previously from a cross between Large White and Meishan breeds but where selection against color had not been made.

### Splice Test

The quantification of wild type *KIT* copies versus copies with the splice mutation was done as previously described (Pielberg et al. 2002).

### Copy Number Test

#### PCR Amplification

Fragments of 189 bp from the duplicated and nonduplicated copies of *KIT* were amplified using one shared forward PCR primer, KITBPF 5'-caactatgctacatccaggc and two specific 5'-biotinylated reverse primers, KIT1BPR 5'-GTGGCTCTTTGAGGTCAG and KIT2BPR 5'-CTGTCTCCATGGTTTGCC. For the tailed-primer approach, a specific tail-sequence (GTAACCGTTCGTACGAGAATCGCT) was

added to the 5'-ends of KIT1BPR and KIT2BPR primers, and a shared 5'-biotinylated tail primer TailR 5'-CGTTCGTACGAGAATCGCT was also included in the PCR. The 5'-biotinylation was needed to allow capture of the PCR products onto streptavidin-coated paramagnetic beads for the preparation of single-stranded DNA. PCR reactions were carried out in a total volume of 25 µL containing 40 ng genomic DNA, 1.5 mM MgCl<sub>2</sub>, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 200 µM dNTPs, 0.75 U AmpliTaq DNA polymerase (PerkinElmer), and 5 pmoles of the KITBPF, KIT1BPR, and KIT2BPR primers. In case of the tailed-PCR approach, 5 pmoles of primers KITBPF and TailR as well as 0.05 pmoles of primers KIT1BPRtailR and KIT2BPRtailR were used. The BAC clones 953F11 and 1041B3 from the porcine *KIT* region (Giuffra et al. 2002) were used for the optimization of the breakpoint test.

Pyrosequencing was carried out following the manufacturer's guidelines and as described (Pielberg et al. 2002), using the PSQ96 instrument (Pyrosequencing AB). The detection primer KITBPseq (5'-ACATACTGTATGATTCCAA) was designed at the very end of the homologous region of the duplicated and non-duplicated copies, and the result is expressed as the ratio between the signals from the incorporated dCTP and dGTP corresponding to the sequence in the two copies (Fig. 1).

### Invader Assay

High-quality genomic DNA samples were prepared from frozen porcine tissue using the DNeasy<sup>TM</sup> Tissue KIT (QIAGEN). Because a minimum DNA concentration is required to obtain adequate signal from the Invader Reaction, the DNA was quantified using the PicoGreen dsDNA Quantitation KIT (Molecular Probes).

An Invader Biplax assay was designed by Third Wave Technologies to estimate the relative number of *KIT* copies in relation to the single-copy gene *ESR*. Invader and probe oligonucleotides were made for *KIT* and *ESR*, being complementary to known exonic sequences (Table 2). Each quantified DNA sample was diluted with QH<sub>2</sub>O to a uniform concentration and then heated to 95°–100°C for 10 min to denature. Invader reactions (15 µL) were set up in duplicate for each sample using the following reagents with their final concentrations: 1 pmol of each Invader oligo, 10 pmol of each probe, 8 mM MgCl<sub>2</sub>, and 100 ng DNA. These reagents were added to a drydown plate containing 100 ng Cleavase VIII enzyme and 5 pmol of each FRET-labeled probe with final buffer concentrations of 10 mM MOPS pH 7.5, 4% PEG, and 2% glycerol. Each well was then overlaid with 25 µL of mineral oil (Sigma), sealed with an adhesive film, and incubated at 63°C for 4 h using a PE9700 thermocycler (Applied Biosystems). Plates were read using the CytoFluor 4000 Fluorescence Plate Reader, first at 485 nm excitation and 530 nm emission for *ESR* and a second time at 560 nm excitation and 620 nm emission for *KIT*. The signal generated by the *KIT* probe was normalized to the signal produced for the control *ESR* probe, thus allowing a copy number call to be made in regards to *KIT*. Standardizing the results to control samples, which were tested on every plate, also allowed the comparison of samples tested on separate microtiter plates.

## ACKNOWLEDGMENTS

We thank Anders Alderborn (Pyrosequencing AB) for valuable discussions on the pyrosequencing technology, and Glen Donald

**Table 2.** Oligonucleotide Probes Used for *KIT* Quantification Based on the Invader Assay

Locus	Invader <sup>a</sup>	Probe <sup>a</sup>
<i>KIT</i>	acttggcggccgaaatatcctcttactT	ACGACGCGGAGatggtcgaatcacaagatt
<i>ESR</i>	gcaggctgacagctctctctgttctT	CGCGCCGAGGatcaattgtgactggtt

<sup>a</sup>The uppercase base in the Invader is the invading base, and the uppercase bases in the probes are the probe arms.

and members of Third Wave Technologies for their help in the development of the Invader Assay system. The study was supported by the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## REFERENCES

- Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., and Eichler, E.E. 2002. Recent segmental duplications in the human genome. *Science* **297**: 1003–1007.
- Bernstein, A., Chabot, B., Dubreuil, P., Reith, A., Nocka, K., Majumder, S., Ray, P., and Besmer, P. 1990. The mouse *W/c-KIT* locus. *Ciba Found. Symp.* **148**: 158–172.
- Eichler, E.E. 2001. Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet.* **17**: 661–669.
- Giuffra, E., Evans, G., Törnsten, A., Wales, R., Day, A., Looft, H., Plastow, G., and Andersson, L. 1999. The Belt mutation in pigs is an allele at the Dominant white (*I/KIT*) locus. *Mamm. Genome* **10**: 1132–1136.
- Giuffra, E., Törnsten, A., Marklund, S., Bongcam-Rudloff, E., Chardon, P., Kijas, J.M.H., Anderson, S.L., Archibald, A.L., and Andersson, L. 2002. A large duplication associated with dominant white color in pigs originated by homologous recombination between LINE elements flanking *KIT*. *Mamm. Genome* **13**: 569–577.
- Hall, J.G., Eis, P.S., Law, S.M., Reynaldo, L.P., Prudent, J.R., Marshall, D.J., Allawi, H.T., Mast, A.L., Dahlberg, J.E., Kwiatkowski, R.W., et al. 2000. Sensitive detection of DNA polymorphisms by the serial invasive signal amplification reaction. *Proc. Natl. Acad. Sci.* **97**: 8272–8277.
- Jackson, I.J. 1994. Molecular and developmental genetics of mouse coat color. *Annu. Rev. Genet.* **28**: 189–217.
- Johansson Moller, M., Chaudhary, R., Hellmén, E., Høyheim, B., Chowdhary, B., and Andersson, L. 1996. Pigs with the dominant white coat color phenotype carry a duplication of the *KIT* gene encoding the mast/stem cell growth factor receptor. *Mamm. Genome* **7**: 822–830.
- Kijas, J.M.H., Wales, R., Törnsten, A., Chardon, P., Moller, M., and Andersson, L. 1998. Melanocortin receptor 1 (*MC1R*) mutations and coat color in pigs. *Genetics* **150**: 1177–1185.
- Kijas, J.M.H., Moller, M., Plastow, G., and Andersson, L. 2001. A frameshift mutation in *MC1R* and a high frequency of somatic reversions cause black spotting in pigs. *Genetics* **158**: 779–785.
- Koppens, P.F., Hoogenboezem, T., and Degenhart, H.J. 2002. Duplication of the *CYP21A2* gene complicates mutation analysis of steroid 21-hydroxylase deficiency: Characteristics of three unusual haplotypes. *Hum. Genet.* **111**: 405–410.
- Lövlie, R., Daly, A.K., Molven, A., Idle, J.R., and Steen, V.M. 1996. Ultrarapid metabolizers of debrisoquine: Characterization and PCR-based detection of alleles with duplication of the *CYP2D6* gene. *FEBS Lett.* **392**: 30–34.
- Lundqvist, E., Johansson, I., and Ingelman-Sundberg, M. 1999. Genetic mechanisms for duplication and multiduplication of the human *CYP2D6* gene and methods for detection of duplicated *CYP2D6* genes. *Gene* **226**: 327–338.
- Lyamichev, V.I., Kaiser, M.W., Lyamicheva, N.E., Vologodskii, A.V., Hall, J.G., Ma, W.P., Allawi, H.T., and Neri, B.P. 2000. Experimental and theoretical analysis of the invasive signal amplification reaction. *Biochemistry* **39**: 9523–9532.
- Lynch, M. 2002. Genomics. Gene duplication and evolution. *Science* **297**: 945–947.
- Marklund, S., Kijas, J., Rodriguez-Martinez, H., Rönstrand, L., Funa, K., Moller, M., Lange, D., Edfors-Lilja I., and Andersson L. 1998. Molecular basis for the Dominant white phenotype in the domestic pig. *Genome Res.* **8**: 826–833.
- Nathans, J. 1999. The evolution and physiology of human color vision: Insights from molecular genetic studies of visual pigments. *Neuron* **24**: 299–312.
- Neve, B., Froguel, P., Corset, L., Vaillant, E., Vatin, V., and Boutin, P. 2002. Rapid SNP allele frequency determination in genomic DNA pools by pyrosequencing. *Biotechniques* **32**: 1138–1142.
- Neville, M., Selzer, R., Aizenstein, B., Maguire, M., Hogan, K., Walton, R., Welsh, K., Neri, B., and de Arruda, M. 2002. Characterization of cytochrome P450 2D6 alleles using the Invader system. *Biotechniques* **32**: S34–S43.
- Ohta, T. 1990. How gene families evolve. *Theor. Popul. Biol.* **37**: 213–219.
- Olivier, M., Chuang, L.M., Chang, M.S., Chen, Y.T., Pei, D., Ranade, K., de Witte, A., Allen, J., Tran, N., Curb, D., et al. 2002. High-throughput genotyping of single nucleotide polymorphisms using new biplex invader technology. *Nucleic Acids Res.* **30**: e53.
- Pielberg, G., Olsson, C., Syvänen, A.-C., and Andersson, L. 2002. Unexpectedly high allelic diversity at the *KIT* locus causing dominant white color in the domestic pig. *Genetics* **160**: 305–311.
- Townson, J.R., Barcellos, L.F., and Nibbs, R.J. 2002. Gene copy number regulates the production of the human chemokine CCL3-L1. *Eur. J. Immunol.* **32**: 3016–3026.
- Wasson, J., Skolnick, G., Love-Gregory, L., and Permutt, M.A. 2002. Assessing allele frequencies of single nucleotide polymorphisms in DNA pools by pyrosequencing technology. *Biotechniques* **32**: 1144–1150.
- Xu, C., Goodz, S., Sellers, E.M., and Tyndale, R.F. 2002. CYP2A6 genetic variation and potential consequences. *Adv. Drug Deliv. Rev.* **54**: 1245–1256.

Received January 17, 2003; accepted in revised form June 17, 2003.



## A Sensitive Method for Detecting Variation in Copy Numbers of Duplicated Genes

Gerli Pielberg, Andy E. Day, Graham S. Plastow, et al.

*Genome Res.* 2003 13: 2171-2177

Access the most recent version at doi:[10.1101/gr.1188203](https://doi.org/10.1101/gr.1188203)

---

**References** This article cites 25 articles, 7 of which can be accessed free at:  
<http://genome.cshlp.org/content/13/9/2171.full.html#ref-list-1>

### License

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---