Genomic Evolution and Polymorphism: Segmental Duplications and Haplotypes at 108 Regions on 21 Chromosomes

Craig A. McLure, Peter Hinchliffe, Susan Lester, Joseph F. Williamson, John A. Millman, Peter J. Keating, Brent J. Stewart, Roger L. Dawkins

PII: S0888-7543(13)00040-2
DOI: doi: 10.1016/j.ygeno.2013.02.011
Reference: YGENO 8490

To appear in: *Genomics*

Received date: 14 June 2012
Accepted date: 27 February 2013

Please cite this article as: Craig A. McLure, Peter Hinchliffe, Susan Lester, Joseph F. Williamson, John A. Millman, Peter J. Keating, Brent J. Stewart, Roger L. Dawkins, Genomic Evolution and Polymorphism: Segmental Duplications and Haplotypes at 108 Regions on 21 Chromosomes, *Genomics* (2013), doi: 10.1016/j.ygeno.2013.02.011

Genomic Evolution and Polymorphism: Segmental Duplications and Haplotypes at 108 Regions on 21 Chromosomes

Craig A McLure[1,2,3,4], Peter Hinchliffe[1], Susan Lester[1,5], Joseph F Williamson*[,1,2], John A Millman[1,6], Peter J Keating[1,2], Brent J Stewart[1,2], Roger L Dawkins[1,2,7]

[1] C.Y. O'Connor ERADE Village, PO Box 5100, Canning Vale, Western Australia 6155, Australia.

[2] School of Veterinary and Biomedical Sciences, Division of Health Sciences, Murdoch University, Murdoch, Western Australia 6150, Australia.

[3] Department of Veterinary Sciences, University of Melbourne, Victoria 3000, Australia.

[4] Genetic Technologies Limited, Fitzroy, Victoria, 3065, Australia.

[5] Rheumatology Department, The Queen Elizabeth Hospital, Woodville, South Australia, Australia.

[6] TAFEWA Swan Campus, Hayman Road, Bentley, Western Australia 6102, Australia.

[7] Faculty of Medicine and Dentistry, University of Western Australia, Nedlands, Western Australia 6009, Australia.

*Corresponding author. Fax: +618 93971559. Tel: +618 93971556

E-mail address: jfw@cyo.edu.au

Abstract

We describe here extensive, previously unknown, genomic polymorphism in 120 regions, covering 19 autosomes and both sex chromosomes. Each contains duplication within multigene clusters. Of these, 108 are extremely polymorphic with multiple haplotypes.

We used the Genomic Matching Technique (GMT), previously used to characterise the Major Histocompatibility Complex (MHC) and Regulators of Complement Activation (RCA).

This genome-wide extension of this technique enables the examination of many underlying *cis*, *trans* and epistatic interactions responsible for phenotypic differences especially in relation to individuality, evolution and disease susceptibility.

The extent of the diversity could not have been predicted and suggests a new model of primate evolution based on conservation of polymorphism rather than *de novo* mutation.

Keywords

Ancestral Haplotypes; Segmental Duplication; Complex Disease; Gene Copy Number Variation; Evolution; Conserved Polymorphism.

1. Introduction

Over half of the observed genetic variation in humans is clustered within genomic regions containing segmental duplications. Interestingly, these polymorphic regions account for only approximately 5% of the genome and tend to be clustered within distinct genomic blocks [1-7]. The principal aim of the present study is to develop a screening test which prospects for biologically important differences and especially those which underlie disease susceptibility and primate evolution. The secondary aim is to determine whether genome-wide polymorphism is sufficient to account for the individuality of humans.

The Genomic Matching Technique (GMT) was developed as an approach to finding suitable bone marrow donors and recipients. After exhaustive testing, the procedure has proven efficient and reliable in recognising alternative polymorphic sequences (haplotypes) within family studies. Identity by GMT predicts a successful transplant outcome [8, 9]. To our initial surprise, it transpires that haplotyping is achieved by amplifying duplicated sequences flanked by highly conserved priming sites. In a new application of the technique, here referred to as "duplotyping", we ask how much polymorphism exists in regions of known duplication. After designing a primer pair with the potential to amplify linked duplicons, we tested each pair by comparing the amplification products from different subjects. In this way, we were able to estimate the extent of polymorphism within each duplicated region.

Previous testing of genomic regions, such as the MHC and the RCA, has confirmed the utility of this approach. Multiple amplification products reflect duplicons of varied lengths as happens when different insertions and deletions (indels) accumulate in one copy rather than another. These indels have been shown to be characteristic of each haplotype so that length can be used for haplotyping and for duplotyping.

Differences in the amount of product of a given length relate to the number of duplicated sequences of that length. Thus, duplication can be detected even when the duplicons have the same sequence in *cis* and have not yet accumulated indels (homoduplications).

Duplotyping of the human MHC has already demonstrated the importance of duplication in polymorphic blocks and their relevance to complex disease [10]. Clusters of multicopy gene families [11-14] are distributed throughout ~3.5 megabases (Mb) and were found to contain extreme levels of polymorphism including genomic duplications, Gene Copy Number Variations (GCNV)s, retroviral and genomic indels and SNPs [15-17]. Specific combinations of these features, including both coding and non-coding polymorphisms, segregated as nuclear haplotypes through multi-generation families [17]. These haplotypes are precise markers of several hundred kilobases (Kb) of sequence [10]. Their occurrence in unrelated individuals implies conservation over many generations and led to the designation Ancestral Haplotypes (AH). Recombination occurs between rather than within blocks [10, 17-20]. The high polymorphic content and the apparent "freezing" of diverse sequences resulted in these regions being termed "Polymorphic Frozen Blocks" (PFBs) [10, 19].

Following the definition of MHC polymorphism, it became possible to investigate the role of genetic susceptibility to diseases. It transpired that particular AHs are associated with specific diseases [10, 21, 22]. The mechanisms responsible are multifactorial and dependent upon haplospecific interactions of coding and non-coding sequences [23, 24]. AHs provide a means of defining these interactions [10, 17, 20, 25] including epistasis.

The extent and importance of ancestral or extended haplotypes was first demonstrated with the identification of haplospecific copy number variations of the complement gene, C4 within the central MHC [26, 27]. Each AH has a specific copy number, which relates, in turn, to serum concentration and susceptibility to disease [28].

Although first discovered in the MHC, quantal structure of the genome is now recognised as characteristic of the entire genome [1, 29, 30], as is the importance of segmental duplications and GCNV on phenotype [31]. Recently high throughput assays such as SNP and Multiplex Ligation-dependent Probe Amplification (MLPA) [32] have been used to detect differences in copy number but with limited success. The MHC and HapMap experiences show that SNP haplotypes of complex regions are misleading. Genomic duplication and especially GCNVs complicate the assignment of

SNPs and the determination of phase remains ambiguous until the haplotypes have been assigned independently by demonstrating inheritance by family segregation [20, 25, 33].

Accordingly, we have developed the genome wide "Duplotyping" approach in order to discover new haplotypes directly. The approach relies upon the amplification of multiple polymorphic elements located within linked duplicons, avoiding the risk of inferring haplotypes from independent SNPs and microsatellites. Each of the duplicons has evolved independently from an ancestral sequence. It follows that the specific combinations of duplicons define informative haplotypes efficiently.

Each test requires a single PCR, making the "Duplotyping" approach an excellent cost-effective and informative alternative to direct sequencing of multiple individuals. The utility of the technique has been demonstrated over decades of clinical practice [34-38]. Matching GMT profiles of donors and recipients predicts a successful bone marrow transplant [8, 9].

GMT has also defined haplotypes in the canine MHC [39], the human RCA [40] and the zebrafish orthologue of human Mannose binding lectin (*MBL2*) [41]. Recently, Lester and colleagues demonstrated an epistatic interaction between the RCA alpha block haplotypes and the MHC in Primary Sjögren's Syndrome [24].

Here we extend the approach to 80 genomic blocks and reveal previously unknown haplospecific polymorphism in humans and in syntenic clusters of other species.

2. Results

2.1. Quantitation and characterization of amplification products

2.1.1. Haplotype analysis in families

An example of the amplification profile is shown in Table 1. In this case, primer pair CYO_5_2 was used to amplify samples from 17 members of a well studied 3 generation CEPH (Centre d'Etude du Polymorphisme Humain) family used to assign individual haplotypes and the resulting composite genotypes throughout the genome.

The raw results (shown in supplementary Figure 1) are tabulated using an internationally verified and reproducible scoring system, which has been proven to reflect copy number [39, 41]. This system allows detection of qualitative and quantitative differences in the amplification products and therefore a precise estimate of polymorphism.

The direct contribution of each haplotype is revealed by comparing the members between and within generations and by demonstrating unequivocal segregation of inheritance.

As shown in Table 1, the grandparents (I1, I1a, I2, I2a) are designated *ab, cd, ef* and *gh* respectively. Their children are *ac* for the father (II1) and e*g* for the mother (II1a). By inspection of the patterns, it is possible to determine which products are attributable to each haplotype. To confirm these assignments, the patterns in the third generation are examined and the haplotypes are assigned.

The results shown in Table 1 are unequivocal because the family has three of the four possible genotypes (*ag, ae, ce, cg)* in the third generation and each has a different pattern as summarised in Figure 1.

Consider, initially, those products in the 242 to 331 basepair range. There are 9 different products within this range including, in this particular case, P2 which is the same length as the 242bp marker resulting in a score of 7 or 8 rather than 3. All scores of 7, 8 or 9 are explained by double doses. Thus, there are 4 products in each subject. Two of these 4 can be assigned to one haplotype and the remaining 2 to the second. For example, every subject with *a* has P4 and P8 whereas *e* has P4 and P11. All with *c* have a score of 7 or 8 for P2 at 242bp and a score of 4, 5, 6 for P5. The *g* haplotype carries P4+P6. The *b, d, f* and *h* haplotypes must be P4+P5, P4+P10, P4+P6 and P3+P9 respectively. Although some products are shared, each haplotype has a unique combination with the single exception of *f* and *g* which share P4+P6. Such sharing is expected given that Ancestral Haplotypes are inherited over many generations.

The unequivocal segregation of products within an informative 3 generation family indicates that the duplicated sequences can be regarded as two polymorphic loci, designated Short and Long, which are closely linked. The resulting haplotypes are inherited faithfully without intervening recombination. The patterns suggest that S and L arose by duplication and that the alleles at each arose by subsequent insertion and deletion (indels).

Similar results were obtained when the same primer pair was tested on other 3 generation families.

2.1.2. Value of secondary interactions between haplotypes.

An individual's profile is due to the amplification of duplicons on the paternal and maternal haplotypes plus any interaction between these primary products. Such interactions can be confounding. For this reason, it is essential to identify those amplicons which are generated directly from the haplotypes and are therefore heritable. Once these are identified, the secondary interactions, as shown on non-denaturing gels, become useful since they define the genotypes or combinations of haplotypes.

Consider, now, the higher molecular weight products shown in Table 1 and supplementary Figure 1. In general, these products relate not to individual haplotypes but to combinations or genotypes. Note, for example, that the *eg* and the *ef* heteroduplexes are very different with either P33 or P41+P 44 +P 54 respectively even though they share haplotype *e* and each of *g* and *f* has P4+ P6. This must mean that there are different secondary interactions between *e* and *g* and between *e* and *f.* In this way, the higher molecular weight heteroduplexes add discrimination and imply polymorphism within the sequences between the priming sites. Note, this additional information has proven to be reproducible although not yet explicable structurally.

2.2. Extent of polymorphism

As a practical screen for the amount of polymorphism, each primer pair was tested against an international panel of typing cells selected to provide a snap shot of human diversity [42]. As one example, primer pair CYO_5_2, which was characterised in families as above, produces complex and

informative patterns. The results are tabulated in Table 2. As in the 1362 family, there are 4 lower molecular weight products (P1 to P16) in most subjects indicating again that each of the 2 duplicons behaves as a polymorphic locus (see also Figure 1). Those with high scores reflecting double doses are deduced to be homozygous at that locus. As expected given the genetic diversity of the panel, there are more products than in family 1362. Those in the range of P5, P6, P7, P8 relate to either locus implying further polymorphism of indels.

The subjects in lanes 14, 15 and 19 appear to have 3 rather than 4 products and raise some interesting possibilities including copy number variation. In keeping with other genomic regions such as C4 on 6p21-22, rare haplotypes could have 1 or 3 rather than the expected 2 loci and some duplications may be homoduplications (identical in *cis*).

Remarkably, for a single primer pair, these haplotype markers create unique patterns in all 30 subjects. There are only two homozygotes, and we estimate that there are more than 30 haplotypes present in this panel. More haplotypes would be expected in the population and therefore there are likely to be at least 50 haplotypes and 2500 genotypes in the population represented by the panel.

This diversity is even greater when the higher molecular weight markers are considered. Note that some individuals have more than 10 heteroduplexes. Others, such as 14, have none, in keeping with a reduction in copy number as postulated above. In Table 3, the results are rearranged to demonstrate the "arrow head" effect. When the primary products are heterozygous and of similar and intermediate length, such as P4, P5, P6 or P7, the number of secondary products increases. Contrariwise, there are fewer and weaker heteroduplexes when the primary products are of very different lengths. This phenomenon increases the discriminatory power of the assay.

By extrapolation of the results in Table 3, just hundreds of basepairs at this single genomic location have the potential to distinguish between the individuals of multiple populations.

2.3. Polymorphic indices

To permit ranking of the degree of polymorphism, we compared various indices such as the total number of products in the panel, the maximal number in any subject, the proportion with unique patterns and the frequency of heterozygotes. All correlated approximately and could be used but the total number of products - including heteroduplexes - was selected as the most complete index for comparing primers on a particular panel. Thus, the score for CYO_5_2 was 66 as shown in Table 3.

2.4. Degrees of polymorphism

The results for 120 primer pairs are shown in Table 4. An adjusted score was used to facilitate multiple comparisons. Note that CYO_5_2 described above is intermediate in ranking; 24 other primer pairs yielded more polymorphism. Many individual subjects had more than 20 products. Some of these polymorphic regions were already known. Note for example 6p22 (MHC)and 1q21(HFE2). On the other hand, many regions identified here were not previously known to be polymorphic and certainly not to the extreme degree revealed in the present study. Interestingly, there was a broad spread across the chromosomes.

2.5. Evolution of syntenic polymorphism

The "ZOO" panel reveals that primer pairs designed to amplify human duplicons also amplify sites in other species. The scores shown in Table 5 are semi-quantitative and indicative only. In the first row, the results against a subset of humans are approximately correlated with the scores shown in Table 4. For example, all with a score of >10 in Table 5 were in the top 34 of the ranking shown in Table 4.

Interestingly, some primer pairs identify extreme polymorphism even when species are separated by hundreds of millions of years. In some cases, such as CYO_9_4, (Surfeit gene cluster) there appears to have been more or less progressive accumulation in vertebrates. CYO_8_1 (Myomesin 2) is more polymorphic in birds than primates.

Some primer pairs are quite selective, as for example CYO_6_3 in the mouse and CYO_22_3 and 22_4 in humans and chimps. Others, such as CYO_10_4 (Supervillin, MAP3K8, Lysozyme like 1 and 2), have been remarkably conserved but in some cases there is apparent drop-out as with CYO_9_3 (interferon type 1cluster) in rodents. Three primer pairs amplified humans and chimps but not orang-utan.

2.6. Significance of duplication and polymorphism

An important aim of the present study was to identify genomic polymorphism of potential relevance to evolution and disease. We therefore asked whether any of the primer pairs might be prioritized by mapping to genomic regions known to influence susceptibility to disease. An example of the approach is shown in Figure 2. The density of disease associations varies greatly along the 10 Mb region selected. The major peak is close to the extreme polymorphism detected by CYO_1_11. Thus, duplotyping defines the specific diseases to be investigated using CYO_1_11.

3. Discussion

3.1. Testing the approach

The first aim of the current study was to define a simple strategy for discovery of genomic polymorphism as a prerequisite for explaining genetic susceptibility to disease. Although GMT was developed for highly polymorphic and clinically relevant regions like the MHC [36, 40, 43], we now show successful extension to the entire human nuclear genome. The approach has identified 120 regions on 21 chromosomes and at least 80 genomic blocks. Some regions are extremely polymorphic; 108 show obvious differences including length, content and copy number, which together define heritable haplotypes. The true magnitude of the diversity is only just becoming apparent, emphasising the need for multicentre studies to explore the relevance of thousands of haplotypes to susceptibility of thousands of diseases.

3.2. Evolution and individuality

The second aim was to estimate whether the degree of genomic diversity is sufficient to explain evolution and individuality. A central tenet of Darwinian theory is that variation is the substrate upon which natural selection operates. Over the past 100 years, most have assumed that sufficient variation could only be possible through ongoing, so-called random, mutation due to, according to general belief, errors in copying DNA. One major weakness of this model is that the variation or, in current terminology, the polymorphism, must be heritable to explain how selected characteristics become inherited and therefore defining for a particular species or population. Mutation, even if initially *de novo* and random, would have to become conserved.

3.3. Heritability and extent

The present results reveal that polymorphism is both heritable and plentiful suggesting that ongoing mutation may not be necessary. Indeed, the polymorphism revealed here must be protected from, or immune to, mutation since it has accumulated and been retained over millions of years.

The degree of polymorphism can be modelled with some minimal assumptions. Assume that there are 100 polymorphic frozen blocks in the human genome. Assume also that there are 100 haplotypes at each block. Since a given individual has paternal and maternal haplotypes at each block there are 100 x 100 possible genotypes at each block. Since meiotic recombination occurs between blocks each of the possible genotypes at each block will be associated randomly with each at other blocks. In a given individual with there are 10^{400} possibilities. Clearly, there are more than sufficient possible genotypes to account for the individuality of the 8 billion humans alive currently and also of all of their ancestors. Contrariwise, to account for human individuality, it is only necessary to postulate that are tens of independently segregating ancestral haplotypes at tens of blocks. Clearly there is more polymorphism than has been appreciated previously.

An attraction of heritable polymorphism is that it also accounts for ancestry. Meiotic recombination shuffles pre-existing polymorphism so as to create differences between siblings without compromising the inheritance of benefits accumulated in previous generations. Thus, given

the vast diversity revealed by the present study, "anamnestic evolution" achieves the twin benefits of individuality and inherited advantage. Popular models of Darwinian evolution based on natural selection of random mutations can be revised given the unexpected degree of inherited, conserved polymorphism now demonstrated.

3.4. Conservation of polymorphism

The concept of conserving polymorphism is far from new [44]. There have been several misunderstandings based, no doubt, on the belief that sequence differences are due to copying errors and therefore have no inherent value until selected. In clinical genetics, unimportant differences are "just polymorphisms". In coding regions, third base "changes" are thought of as random mutations. By contrast, immunogenetics for transplant matching has shown that many sequences are inherited faithfully over many generations and contribute to the description of haplotypes which are critical to biological outcomes. Many polymorphisms are actually trans-species [45] in that they have survived speciation events and can therefore contribute accumulated benefits. In a study of different breeds of cattle, we began the process of classifying differences into the many which are conserved and the few which may represent relatively recent mutation [46].

Conservation of polymorphism must be important in understanding evolution. Differences between individuals must be heritable to be consequential. Patently, physical and molecular characteristics which define an individual are at least largely heritable and traceable to recent, if not remote, ancestors. The differences between dizygotic twins and siblings represent the meiotic shuffling of inherited features which, in the absence of shuffling, are directly responsible for the similarity of monozygotic twins.

It follows that the degree of heritable, conserved polymorphism may be far greater than demonstrated hitherto and of fundamental importance. The difficulty in the past has been that the only valid approach is to sequence whole haplotypes after finding those which are immune to meiotic recombination. Effectively this has meant undertaking exhaustive 3 generation family studies of each polymorphic frozen block within the genome. The power of this approach has been

demonstrated using CEPH families but the magnitude of the effort on a genome-wide basis has been daunting. Even with advances in sequencing technology there is a need for a means of targeting the most informative regions and then their conserved haplotypes.

Duplotyping provides a practical approach and has proven effective in revealing at least 108 promising targets. More interestingly, it has also shown that the degree of polymorphism is greater than expected. Indeed, it is no longer surprising to envisage that each individual, other than monozygotic twins, has a unique and yet heritable DNA sequence.

3.5. Speciation and disease

The ZOO panel used here shows remarkable conservation of priming sites *inter alia* but also species specific differences in architecture, including indels and deduced copy number. Although sequences may be trans-species rather than de novo, it is clear by comparing different species that rearrangements of polymorphic elements must occur. In a previous study we concluded that retroviral sequences were important in speciation events [10, 47] and that duplication, insertion and deletion contribute to the creation of new genotypes utilising basically conserved sequences.

Having defined new polymorphic blocks and some of their haplotypes, it is now possible to ask questions concerning their importance in evolution and diverse diseases. This investigation is proceeding in pilot form but requires multicentre attention.

4. Materials and Methods

The stepwise approach can be summarised:

4.1. Identifying the duplicated Segments

Large genomic segments, known to contain duplications [5], were downloaded from the NCBI human genome assembly, build 35.1 (http://www.ncbi.nlm.nih.gov/projects/genome/guide/human/index.shtml) and examined using Accelrys Gene 2.5 (http://accelrys.com/solutions/science/biosciences/). Dot plot analyses, DOTTER [48] and Gepard (http://www.helmholtz-muenchen.de/en/mips/services/analysis-

tools/gepard/index.html) were used to determine the limits of the duplicated segments. Duplicons with a total length greater than 10 Kb, inclusive of retroelements, were recorded and analysed in further detail.

4.2. Element Discovery and Examination

Duplicons were examined for small (50-1000 nucleotide) horizontal and vertical shifts (indels) in the conserved diagonal line of consensus. Once identified, the elements were examined at the sequence level. The elements targeted ideally contain imperfect, repetitive units that exhibit some additional form of geometric complexity. The rationale is that complex, imperfect elements are less prone to slippage and mutation and therefore more stable than simple, perfect repeats such as dinucleotide microsatellites. Since the aim is to identify polymorphism indicative of AHs, stability throughout human evolution is essential.

4.3. Conservation of Flanking Regions

Regions flanking elements must display sufficient conservation between duplicons to allow binding and amplification of each copy by a single PCR primer pair and be close enough to allow robust amplification. Occasional SNPs are permitted within the primer sites, except within the last 5 bases of the 3` end.

4.4. Unlinked Amplification

The approach relies upon the specific amplification of linked duplicons. To avoid unlinked amplification, elements are screened for retroviral sequence and paralogous copies. Most of the problems in this regard were avoided by systematic *in silico* modelling as described below.

4.5. Retroviral Elements

Sequences were examined using Repeat Masker (http://repeatmasker.org/cgi-bin/WEBRepeatMasker). Based on the results, elements were either accepted or rejected. Those that contain no evidence of retroviral sequence or have a well dispersed retroviral sequence, such as

a LINE, at either the 5` or 3` end were included. Elements containing short, high frequency retroelements such as *Alu*s, were rejected.

4.6. Paralogous Copies

Elements were examined using BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat) to identify any paralogous copies.

4.7. Primer Design and Evaluation

Following the examination for retroelements and paralogous copies, sequences were submitted to Primer 3 (http://frodo.wi.mit.edu/primer3/input.htm), with an optimal primer annealing temperature set to ~60 degrees Celsius. Primer combinations selected by Primer 3 were compared to the alignment of the elements. The primers that were most likely to result in binding and amplification of the intended duplicons were selected and analysed further. As a final examination, primer pairs are submitted to BLAT to exclude paralogous amplification. Primers were manufactured by Sigma-Genosys Oligos (http://www.sigmaaldrich.com/life-science/custom-oligos.html).

4.8. Polymorphism analysis

Samples for this study include: a) 30 ethnically diverse and well defined samples from the International Histocompatibility Workshop Group (http://www.ihwg.org/order/blcl.html) [42] and b) three families (CEPH/Utah Pedigree 1362, CEPH/Amish Pedigree 884 and Venezuelan Pedigree 104) from Coriell Cell Repositories (http://ccr.coriell.org/).

4.9. Polymerase Chain Reaction:

PCRs were performed in a 96-well Palm Cycler (Corbett Research) in 20 microliter volumes using conditions previously described [40]. Optimal primer annealing temperatures were defined prior to interrogating the 4AOH panel and ranged between 52 and 62 degrees Celsius (Supplementary Table 1).

4.10. Detection of amplicons and haplotypes:

The separation and detection of haplotypes was performed using a Corbett Research GS-3000 automated gel analysis system as previously described [40]. A pUC 19 (Fisher Biotech) molecular weight ladder was included. Amplicons were numbered according to their relative migration during non-denaturing electrophoretic separation (Fig. 1). Relative intensities were tabulated using a range from 1-9, where 1 is negative, 2 is equivocal and 3-9 are positive and relative (Table 1).

4.11. Examination of the Approach to Syntenic Clusters

The CYO DNA zoo panel consists of 4 humans (*Homo sapiens*), 2 chimpanzees (*Pan troglodytes*), 1 orang-utan (*Pongo pygmaeus*), 1 rhesus monkey (*Macaca mulatta*), 3 cows (*Bos taurus*), 3 sheep (*Ovis aries*), 5 horses (*Equus caballus*), 5 dogs (*Canis familiaris*), 1 mouse (*Mus musculus*), 1 rat (*Rattus norvegicus*), 1 snake (*Pseudonaja affinis*), 1 chicken (*Gallus gallus*), 1 budgerigar (*Melopsittacus undulatus*), 1 axolotl (*Ambystoma mexicanum*), 1 zebrafish (*Danio rerio*), 1 marron (*Cherax tenuimanus*) and 1 honeybee (*Apis mellifera*) (Table 5). To accommodate variations in primer-binding site sequences, annealing temperatures were reduced by 5 degrees Celsius. All other conditions were as previously described. Results of the analysis are reported in Table 5 and show the maximum number of amplicons observed per individual within each species.

4.12. Genome Wide Identification of Critical Regions: Phenotype Analysis

Figure 2 is a graphical representation of the aggregate disease frequency per Mb of each chromosome and was derived from examination of the OMIM and Phenotype resources at NCBI (NCBI Build 35.1); (http://www.ncbi.nlm.nih.gov/).

The number of genes with OMIM links, per Mb of each chromosome, was calculated. Similar analysis of the Phenotype data was performed, including the multiple records observed at each locus. Totals per Mb region of each chromosome were tabulated. OMIM and Phenotype results were multiplied and recorded for each Mb. The average for each chromosome was subtracted from this value and the results smoothed by comparing to neighbouring regions. Negative values were

removed and results plotted as percentage of value at each Mb compared to the maximum value on the chromosome. Results of this analysis can be seen in Figs. 2.

Abbreviations:

AH – Ancestral Haplotype

CEPH -Centre d'Etude du Polymorphisme Humain

GCNV – Gene Copy Number Variation

GMT – Genomic Matching Technique

MHC – Major Histocompatibility Complex

PFB – Polymorphic Frozen Block

RCA – Regulators of Complement Activation

Competing Interest

Collectively, the authors associated with the C.Y. O'Connor ERADE Village have an interest in Genetic Technologies Ltd.

Authors' contributions

CM jointly conceived the study with RD and carried out the genomic analysis, molecular genetic studies, database design and drafted the manuscript with the active participation of RD, JW, and SL. PH designed and developed the database of phenotypic variants and contributed to the drafting of the manuscript. SL participated in the design of the study, performed the statistical analysis and contributed to the drafting of the manuscript. JW participated in the genomic analysis and molecular genetics studies. JM participated in the design of the study and the statistical analysis. PK

participated in the genomic analysis and drafting the manuscript. BJS participated in the design of the study, the statistical analysis and the drafting of the manuscript. RD jointly conceived the study and participated in its design and coordination and is responsible for the final draft of the manuscript.

Acknowledgements

The research was supported by the C Y O'Connor Village Foundation and Genetic Technologies Ltd., Fitzroy, Victoria 3065, Australia. Software is available on request (admin@cyo.edu.au). The authors would like to thank Dean Male and Patrick Carnegie for their comments on the manuscript and Sally Lloyd and Natalie Jacobsen for analyses, tables and valuable suggestions.

References

[1] S.B. Gabriel, S.F. Schaffner, H. Nguyen, J.M. Moore, J. Roy, B. Blumensteil, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S.N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E.S. Lander, M.J. Daly, D. Altshuler, The Structure of Haplotype Blocks in the Human Genome, Science, 296 (2002) 2225-2229.

[2] M.S. Phillips, R. Lawrence, R. Sachidanandam, A.P. Morris, D.J. Balding, M.A. Donaldson, J.F. Studebaker, W.M. Ankener, S.V. Alfisi, F.S. Kuo, A.L. Carmisa, V. Pazorov, K.E. Scott, B.J. Carey, J. Faith, G. Katari, H.A. Bhatti, J.M. Cry, V. Derohannessian, C. Elosua, A.M. Forman, N.M. Grecco, C.R. Hock, J.M. Kuebler, J.A. Lathrop, M.A. Mockler, E.P. Nachtman, S.L. Restine, S.A. Varde, M.J. Hozza, C.A. Gelfand, J. Broxholme, G.R. Abecasis, M.T. Boyce-Jacino, L.R. Cardon, Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots, Nature Genetics, 33 (2003) 382-387.

[3] K. Tokunaga, G. Saueracker, P.H. Kay, F.T. Christiansen, R. Anand, R.L. Dawkins, Extensive deletions and insertions in different MHC supratypes detected by pulsed field gel electrophoresis, Journal of Experimental Medicine, 168 (1988) 933-940.

[4] T. Reichhardt, Patent on gene fragment sends researchers a mixed message...as Germany hesitates over Brussels directive, Nature, 396 (1998) 499.

[5] J.P. Himanen, M. Henkemeyer, D.B. Nikolov, Crystal structure of the ligand-binding domain of the receptor tyrosine kinase EphB2, Nature, 396 (1998) 486-491.

[6] W.J. Zhang, M.A. Degli-Esposti, T.J. Cobain, P.U. Cameron, F.T. Christiansen, R.L. Dawkins, Differences in gene copy number carried by different MHC ancestral haplotypes. Quantitation after physical separation of haplotypes by pulsed field gel electrophoresis, Journal of Experimental Medicine, 171 (1990) 2101-2114.

[7] K. Howard, Developmental biology. Attractive genetics, Nature, 396 (1998) 406-407.

[8] Impact of antiretroviral therapy on tuberculosis incidence among HIV-positive patients in high-income countries, Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 54 (2012) 1364-1372.

[9] C. Witt, D. Sayer, F. Trimboli, M. Saw, R. Herrmann, P. Cannell, D. Baker, F. Christiansen, Unrelated Donors Selected Prospectively by Block-matching Have Superior Bone Marrow Tranpslant Outcome, Human Immunology, 61 (2000) 85-91.

[10] G.M. Huang, Growing pains of a genomics industry, Nature, 396 (1998) 307.

[11] S. Bahram, M. Bresnahan, D.E. Geraghty, T. Spies, A second lineage of mammalian major histocompatibility complex class I genes., Proceedings of the National Academy of Sciences of the United States of America, 91 (1994) 6259-6263.

[12] S. Gaudieri, C. Leelayuwat, D.C. Townend, J.K. Kulski, R.L. Dawkins, Genomic characterization of the region between HLA-B and TNF: implications for the evolution of multicopy gene families, Journal of Molecular Evolution, 44 (1997) S147-S154.

[13] R.A. Weinberg, Telomeres. Bumps on the road to immortality, Nature, 396 (1998) 23-24.

[14] L. Pichon, G. Carn, P. Bouric, T. Giffon, B. Chauvel, M. Lepourcelet, J. Mosser, J.Y. Legall, V. David, Structural analysis of the HLA-A/HLA-F subregion: precise localization of two new multigene families closely associated with the HLA class I sequences, Genomics, 32 (1996) 236-244.

[15] S. Gaudieri, R.L. Dawkins, K. Habara, J.K. Kulski, T. Gojobori, SNP profile within the Human Major Histocompatibility Complex reveals an extreme and interrupted level of nucleotide diversity, Genome Research, 10 (2000) 1579-1586.

[16] S. Gaudieri, J.K. Kulski, R.L. Dawkins, T. Gojobori, Extensive nucleotide variability within a 370 kb sequence from the central region of the Major Histocompatibility Complex, Gene, 238 (1999) 157-161.

[17] M.A. Degli-Esposti, A.L. Leaver, F.T. Christiansen, C.S. Witt, L.J. Abraham, R.L. Dawkins, Ancestral haplotypes: conserved population MHC haplotypes, Human Immunology, 34 (1992) 242-252.

[18] K.H. Rhee, E.P. Morris, J. Barber, W. Kuhlbrandt, Three-dimensional structure of the plant photosystem II reaction centre at 8 A resolution, Nature, 396 (1998) 283-286.

[19] B. Marshall, C. Leelayuwat, M.A. Degli-Esposti, M. Pinelli, L.J. Abraham, R.L. Dawkins, New major histocompatibility complex genes, Human Immunology, 38 (1993) 24-29.

[20] E.J. Yunis, C.E. Larsen, M. Fernandez-Vina, Z.L. Awdeh, T. Romero, J.A. Hansen, C.A. Alper, Inheritable variable sizes of DNA streches in the human MHC: conserved extended haplotypes and their fragments or blocks, Tissue Antigens, 62 (2003) 1-20.

[21] S. Jenisch, E. Westphal, R.P. Nair, P. Stuart, J.J. Voorhees, E. Christophers, M. Kronke, J.T. Elder, T. Henseler, Linkage disequilibrium analysis of familial psoriasis: identification of multiple disease-associated MHC haplotypes, Tissue Antigens, 53 (1999) 135-146.

[22] M.A. Degli-Esposti, C. Leelayuwat, R.L. Dawkins, Ancestral haplotypes carry haplotypic and haplospecific polymorphisms of BAT1: possible relevance to autoimmune disease, European Journal of Immunogenetics, 19 (1992) 121-127.

[23] B. Gold, J.E. Merriam, J. Zernant, L.S. Hancox, A.J. Taiber, K. Gehrs, K. Cramer, J. Neel, J. Bergeron, G.R. Barile, R.T. Smith, G.S. Hageman, M. Dean, R. Allikmets, S. Chang, L.A. Yannuzzi, J.C. Merriam, I. Barbazetto, L.E. Lerner, S. Russell, J. Hoballah, J. Hageman, H. Stockman, Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration, Nat Genet, 38 (2006) 458-462.

[24] S. Lester, C. McLure, J. Williamson, P. Bardy, M. Rischmueller, R.L. Dawkins, Epistasis between the MHC and the RCA alpha block in primary Sjogren syndrome, Ann Rheum Dis, 67 (2008) 849-854.

[25] C.A. Alper, C.E. Larsen, D.P. Dubey, Z.L. Awdeh, D.A. Fici, E.J. Yunis, The haplotype structure of the human major histocompatibility complex, Hum Immunol, 67 (2006) 73-84.

[26] D. Raum, Z. Awdeh, J. Anderson, L. Strong, J. Granados, L. Teran, E. Giblett, E.J. Yunis, C.A. Alper, Human C4 haplotypes with duplicated C4A or C4B, Am J Hum Genet, 36 (1984) 72-79.

[27] T.P. Evgen'eva, I.V. Semenova, Comparative characteristics of the muscular tissue in bony fishes with different nutrition types, Dokl Biol Sci, 370 (2000) 81-83.

[28] G. Uko, F.T. Christiansen, R.L. Dawkins, P.H. Kay, Low serum C4 concentrations in insulin dependent diabetes mellitus., British Medical Journal, 286 (1983) 1748 - 1749.

[29] M.J. Daly, J.D. Rioux, S.F. Schaffner, T.J. Hudson, E.S. Lander, High-resolution haplotype structure in the human genome, Nature Genetics, 29 (2001) 229-232.

[30] N. Longman-Jacobsen, J.F. Williamson, R.L. Dawkins, S. Gaudieri, In polymorphic genomic regions indels cluster with nucleotide polymorphism: Quantum Genomics, Gene, 312 (2003) 257-261.

[31] J. Sebat, B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, H. Massa, M. Walker, M. Chi, N. Navin, R. Lucito, J. Healy, J. Hicks, K. Ye, A. Reiner, T.C. Gilliam, B. Trask, N. Patterson, A. Zetterberg, M. Wigler, Large-scale copy number polymorphism in the human genome, Science, 305 (2004) 525-528.

[32] P.W. Schouten, A.V. Parisi, Underwater deployment of the polyphenylene oxide dosimeter combined with a neutral density filter to measure long-term solar UVB exposures, J Photochem Photobiol B, 112 (2012) 31-36.

[33] R.R. Wener, F.J. Loupatty, W.E. Schouten, Isolated elevated aspartate aminotransferase: a surprising outcome for clinicians, Neth J Med, 70 (2012) 136-138.

[34] F.T. Christiansen, G. Tay, L.K. Smith, C.S. Witt, E.W. Petersdorf, B. Bradley, R.L. Dawkins, Histocompatibility matching for bone marrow transplantation. Donor-recipient pairs in the 4AOHW cell panel, Human Immunology, 38 (1993) 42-51.

[35] G.K. Tay, C.S. Witt, F.T. Christiansen, D. Charron, D. Baker, R. Herrmann, L.K. Smith, D. Diepeveen, S. Mallal, J. McCluskey, S. Lester, P. Loiseau, H. Teisserenc, J. Chapman, B. Tait, R.L. Dawkins, Matching for MHC haplotypes results in improved survival following unrelated bone marrow transplantation, Bone Marrow Transplantation, 15 (1995) 381-385.

[36] G.K. Tay, C.S. Witt, F.T. Christiansen, J.M. Corbett, R.L. Dawkins, The identification of MHC identical siblings without HLA typing, Experimental hematology, 23 (1995) 1655-1660.

[37] H. Grosse-Wilde, N. Ketheesan, F. Christiansen, H.D. Ottinger, S. Ferencik, G.K. Tay, C.S. Witt, H. Teisserenc, M. Giphart, E.M. Freitas, D. Charron, R.L. Dawkins, The genomic matching technique (GMT): A new tool for selecting unrelated marrow donors, in: D. Charron (Ed.) 12th International Histocompatibility Workshop and Conference, EDK, Paris, France, 1997, pp. 589-591.

[38] N. Ketheesan, S. Gaudieri, C.S. Witt, G.K. Tay, D.C. Townend, F.T. Christiansen, R.L. Dawkins, Reconstruction of the Block Matching Profiles, Human Immunology, 60 (1999) 171-176.

[39] C.A. McLure, P.W. Kesners, S. Lester, D. Male, C. Amadou, J.R. Dawkins, B.J. Stewart, J.F. Williamson, R.L. Dawkins, Haplotyping of the canine MHC without the need for DLA typing, Int J Immunogenet, 32 (2005) 407-411.

[40] C.A. McLure, J.F. Williamson, L.A. Smyth, S. Agrawal, S. Lester, J.A. Millman, P.J. Keating, B.J. Stewart, R.L. Dawkins, Extensive genomic and functional polymorphism of the complement control proteins, Immunogenetics, 57 (2005) 805-815.

[41] F.J. Verdam, P.R. Liedorp, N. Geubbels, R. Schouten, I.M. Janssen, G.H. Koek, J.W. Greve, [EndoBarrier for counteracting obesity and metabolic syndrome], Ned Tijdschr Geneeskd, 156 (2012) A3844.

[42] S.K. Cattley, J.F. Williamson, G.K. Tay, O.P. Martinez, S. Gaudieri, R.L. Dawkins, Further characterization of MHC haplotypes demonstrates conservation telomeric of HLA-A: Update of the 4AOH and 10 IHW cell panels, European Journal of Immunogenetics, 27 (2000) 397-426.

[43] C. McLure, P. Kesners, S. Lester, D. Male, C. Amadou, J. Dawkins, B. Stewart, J. Williamson, R. Dawkins, Haplotyping of the canine MHC without the need for DLA typing, International Journal of Immunogenetics, 32 (2005) 407–411.

[44] G.T. Lezin, K.V. Makarova, V.V. Velikodvorskaia, E.S. Zelentsova, R.R. Kechumian, M.G. Kidwell, E.V. Kunin, M.B. Evgen'ev, [Structure and evolutionary role of the Penelope mobile element in Drosophila species of the virilis group], Mol Biol (Mosk), 35 (2001) 805-815.

[45] J. Klein, Origin of major histocompatibility complex polymorphism: the trans-species hypothesis, Human Immunology, 19 (1987) 155-162.

[46] G.S. Hageman, L.S. Hancox, A.J. Taiber, K.M. Gehrs, D.H. Anderson, L.V. Johnson, M.J. Radeke, D. Kavanagh, A. Richards, J. Atkinson, S. Meri, J. Bergeron, J. Zernant, J. Merriam, B. Gold, R. Allikmets, M. Dean, Extended haplotypes in the complement factor H (CFH) and CFH-related (CFHR) family of genes protect against age-related macular degeneration: characterization, ethnic distribution and evolutionary implications, Ann Med, 38 (2006) 592-604.

[47] J.K. Kulski, S. Gaudieri, R.L. Dawkins, Using Alu J Elements as Molecular Clocks to Trace the Evolutionary Relationships Between Duplicated HLA Class I Genomic Segments, Journal of Molecular Evolution, 50 (2000) 510-519.

[48] E. Sonnhammer, R. Durbin, A dot-matrix program with dynamic threshold control suited for genomic DNA and protien sequence analysis, Gene, 167 (1996).

Table 1 Tabulation and analysis of products from a 3 generation family
CEPH Pedigree 1362

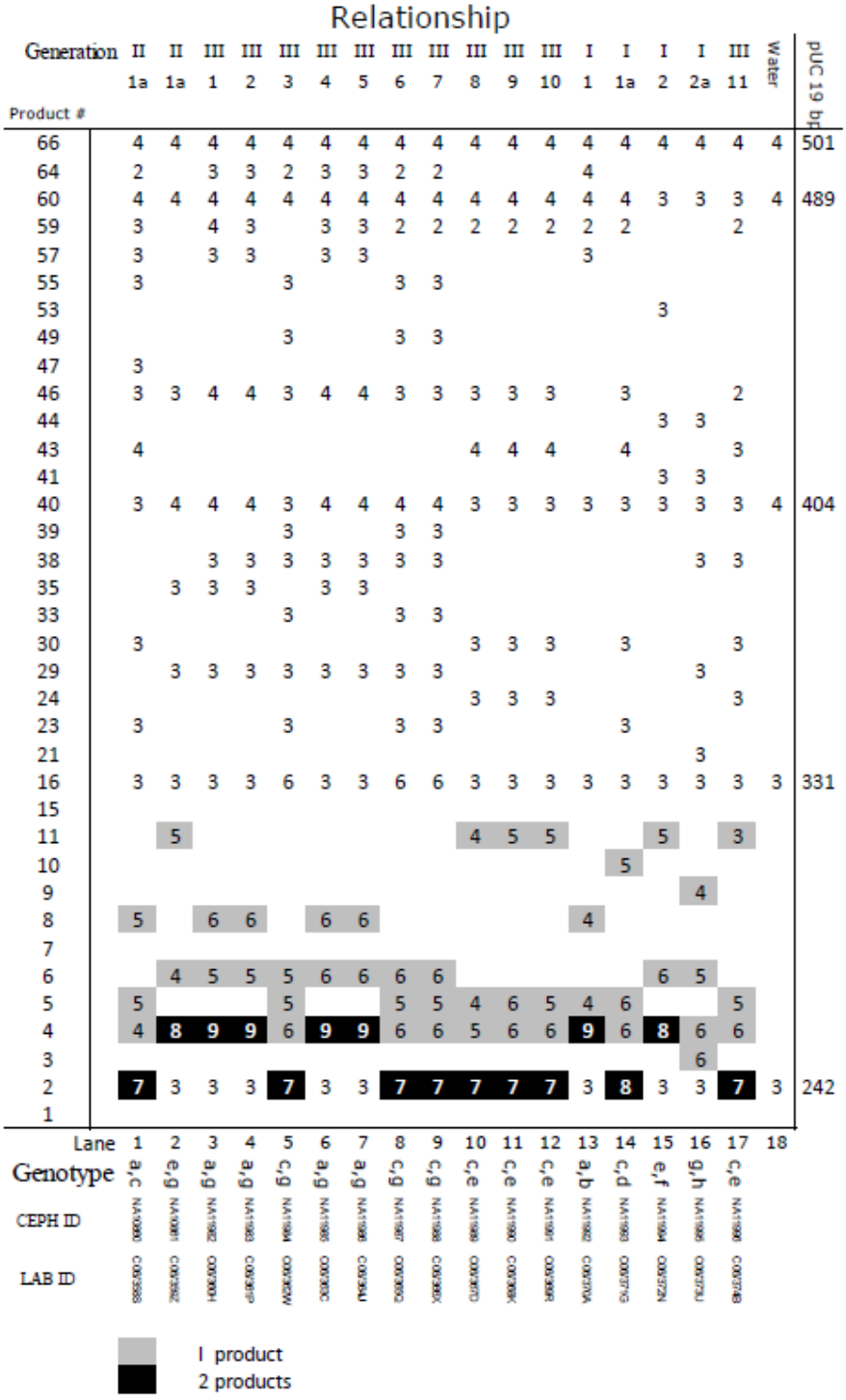


Table 2 Tabulation of panel products CYO_5_2_final

																																pUC 19 bp
16	4	3	3	3	3	3	5	3	3	5	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	331
15								3																				3				
14						3																									3	
13				5																												
12														5		5	5													5		
11				6		5							5								5				5							
10		7							5								5						5			5						
9								5				5						5	7			5	5			6		5				
8		7	7		5		6	6			5						5				5											
7		8	9		6					8	9		6					6		5				6			9	5	8	4		
6				7					5	6		5		9	5	6		7		6				2		5			8			
5					6		8	6	2	6	6	5	6		6	5						8		9	5	6	8	6		7		
4		7	7			9	9	6	8			4	5			5	4		8	4	4		7					5			5	
3				8	7	7									6					5	5	5			9						9	
2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	242
1																		9														
Lane	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
Workshop ID	pUC 19	BEA, PL	DAR, KR	REE, GD	JESTHOM	VAVY	BBM	JVM	TBI	EK	BMG	LWA08	HO1M	EJ028	HOKKAIDO	HARA	NPC 439	NON, L	HAY, BD	LOS41265	FIS, WJ	IHL_AD036	MIN	95042860	96900868	T1827	CRB	LB	JBUSH	KT3		
Lab ID	LB	Q86/7062	R86/10429	R85/15168	R86/12289C	R86/12308	R86/12317	R86/12324	R86/12337	R86/12338	R86/12353	R86/12364R	R86/12367	R86/12379W	U86/4390C	R87/4700C	R89/2518M	R81/7077Y	Q82/1071C	R86/12371U	R82/16138	R82/20603	R82/2177R	R87/5531P	R87/5291G	U86/4291R	U86/4238K	U86/4185Y	U86/4291R	U86/4059T	R86/14082	

Table 3 Tabulation and analysis of heteroduplex products

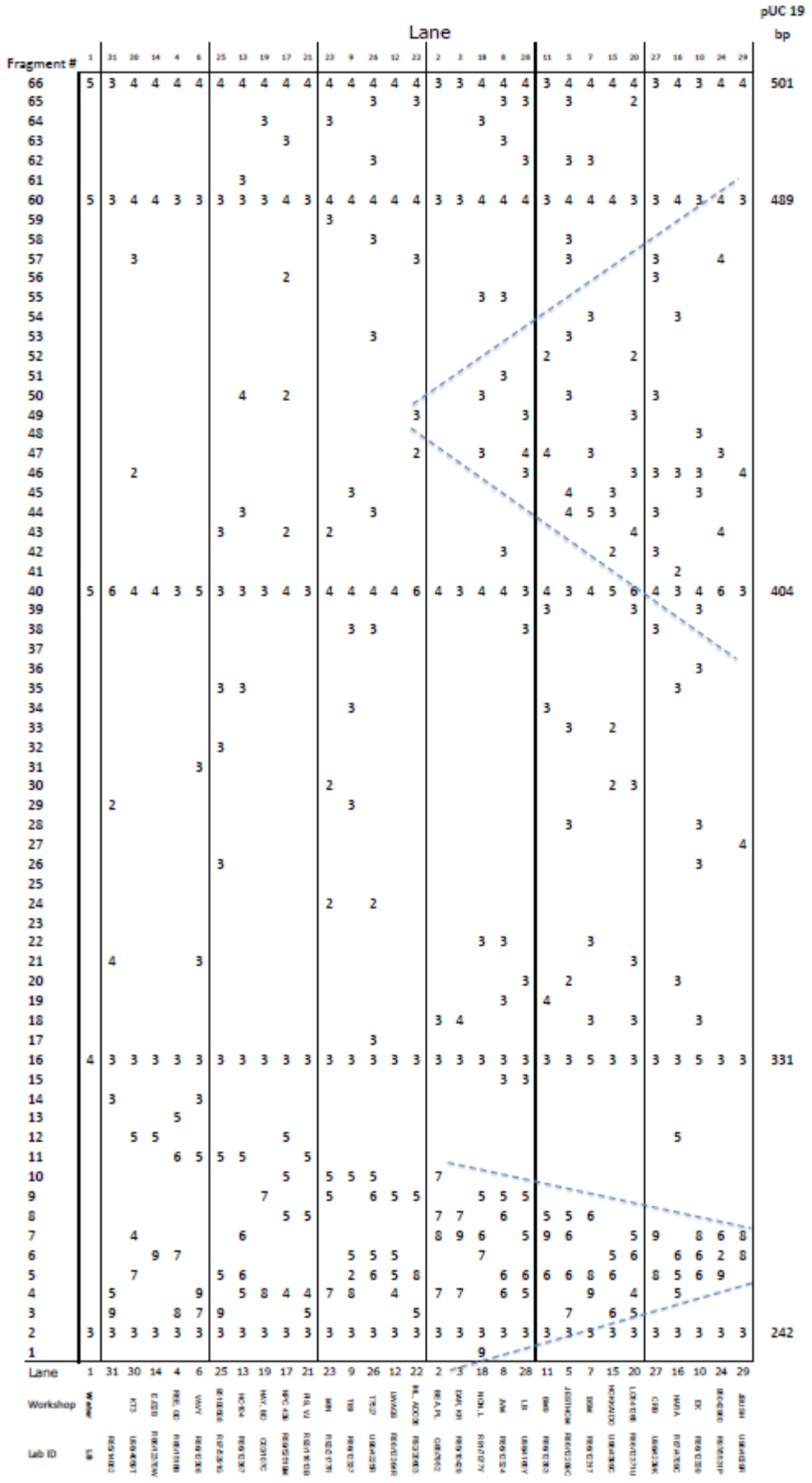


Table 4 Degrees of polymorphism

Name (CYO_)	Region	Total # Products
3_2	3p21-cen to 3q11.2	58
6_5	6p22	57
1_3	1p36.22	55
6_1	6p11.2	52
6_6	6p22	52
22_3	22q11.23	45
8_3	8p23.1	43
10_7	10q22.3	43
1_13	1p36	42
19_1	19q13.2	40
Y_9	Yq11.23	39
10_6	10q22.3	38
1_6	1p21	35
4_3	4q28-q31	35
X_13	Xq28	34
1_11	1q21-q23	33
9_4	9q34	32
10_2	10p11.2	32
22_4	22q13.1	32
Y_11	Yq11.23	31
X_5	Xp11.22	30
5_3	5q35.3	28
15_3	15q21.1	28
2_5	2q12.3-q13	27
1_5	1p22.2	26
5_2	5q21.1	26
12_3	12q24.33	26
15_1	15q11.2	25
17_6	17q12	25
X_3	Xp11.23	25
7_4	7q11.23	24
10_3	10p11.2	24
9_3	9p21-p22	23
12_2	12p11	23
17_3	17q11.2	22
X_8	Xq22.1	22
Y_8	Yq11.22	22
2_1	2p13.1	21
8_5	8p23.1	21
2_8	2q21.1	20
8_4	8p23.1	20
10_4	10p11.2	20
10_5	10q22.3	20
17_5	17q12	19
Y_7	Yq11.2	19
11_1	11p15.4	18
15_4	15q23	18
X_15	Xq28	18
7_2	7p14-p15	17
15_2	15q13.1	17
Y_3	Yp11.2	17
10_1	10p11.2	16
17_4	17q11.2	16
Y_4	Yp11.2	16
1_4	1p36.13	15
7_3	7p11	15
14_2	14q32.33	15
22_2	22q11.21	15
Y_5	Yq11.2	15
4_1	4q13	14

Table 5 Syntenic Clusters

Species	Common Ancestor (MYA)	CYO_8_1	CYO_10_4	CYO_9_4	CYO_9_3	CYO_Y_4	CYO_Y_8	CYO_6_3	CYO_Y_11	CYO_22_4	CYO_5_1	CYO_Y_10	CYO_6_4	CYO_X_14	CYO_17_5	CYO_Y_9	CYO_12_2	CYO_19_1	CYO_3_3	CYO_5_2	CYO_10_3	CYO_10_7	CYO_12_1	CYO_22_3	CYO_15_3	CYO_2_8	CYO_10_6	CYO_X_15	CYO_3_2	CYO_Y_7	CYO_2_5	CYO_Y_5	n
Human	0	8	9	>10	>10	6	5	2	>10	6	6	7	>10	5	5	10	7	9	6	5	>10	>10	5	>10	3	7	9	7	4	5	8	4	4
Chimpanzee	5	5	8	>10	>10	3	6	3	8	4	3	7	7	8	5	7	4	9	4	5	>10	>10	2	8	8	3	4	4	4	3	3	4	2
Orangutan	20	6	7	>10	>10	2	1	0	1	1	2	1	6	4	3	9	4	4	4	1	3	>10	1	0	3	2	3	2	2	1	2	0	1
Rhesus	30-40	1	3	>10	>10	6	5	1	6	1	3	1	>10	8	10	5	5	3	3	3	8	6	3	3	5	2	5	6	2	2	1	0	1
Mouse	90	4	5	8	5	5	1	>10	3	0	0	0	8	>10	7	5	3	5	1	0	0	0	0	0	5	0	0	0	0	0	0	0	1
Rat	90	5	2	<10	3	2	2	1	0	0	1	1	2	3	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Dog	140	4	>10	>10	>10	3	4	0	2	2	0	0	0	2	4	5	5	0	0	2	7	3	3	1	1	1	0	0	0	0	0	2	5
Horse	140	8	>10	>10	>10	5	5	4	0	3	3	1	4	8	4	>10	3	3	2	1	>10	5	1	5	3	1	0	0	1	1	0	1	5
Cow	140	8	8	>10	>10	2	3	3	1	3	2	0	4	8	5	6	1	1	0	0	>10	5	3	1	0	0	1	1	0	0	0	0	3
Sheep	140	7	8	>10	>10	3	4	6	1	1	3	0	6	4	3	2	2	0	1	1	9	3	1	0	0	0	1	0	0	0	0	0	3
Chicken	220	>10	9	7	>10	4	7	2	0	0	1	1	4	4	4	5	1	1	1	2	3	2	0	0	0	0	0	0	0	0	0	0	1
Budgerigar	220	>10	8	9	>10	1	5	4	1	0	0	0	3	6	2	6	4	4	5	1	4	0	0	0	0	0	0	0	0	0	0	0	1
Snake	220	7	6	8	4	0	1	1	3	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Axolotl	350	>10	3	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Zebrafish	365	2	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Honeybee	630	2	5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	1
Marron	630	10	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Fig. 1 Summary of product profiles in a 3 generation family

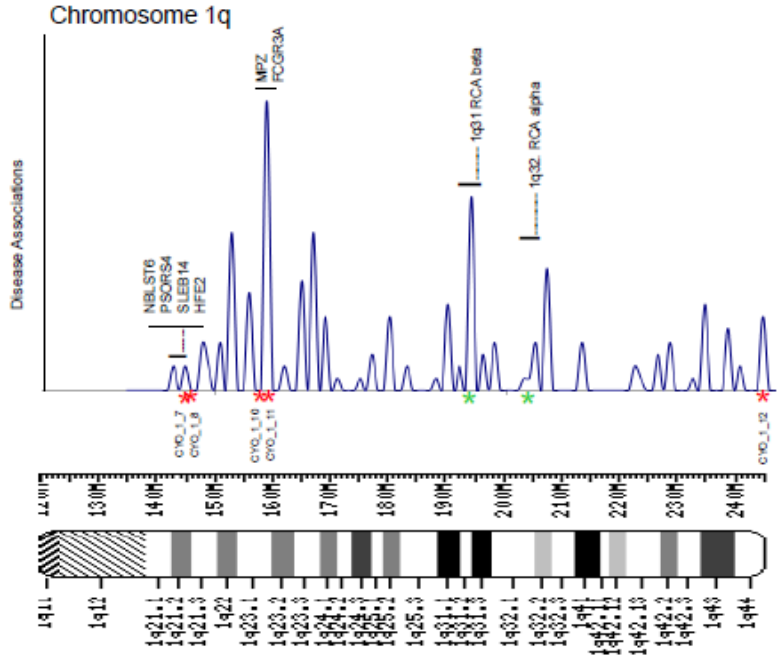
Heteroduplex Products

Fragment #	e g	e f	g h	a g	c g	a b	c e	c d	a c	pUC 19 bp
66										501
64				I	w	I			I	
60										489
59				I		w	w	w	I	
58			I							
57				I		I				
55					I				I	
54		I	I							
49					I					
47									I	
46				I	I		I	I		
44		I	I							
43							I	I	I	
41		I	I							
40										404
39					I					
38	I	I	I	I	I					
35	I			I						
33					I					
30							I	I	I	
29	I		I	I	I					
24							I			
23					I			I	I	
21			I							
16					I					331

Genotype Products

	e g	e f	g h	a g	c g	a b	c e	c d	a c
11	I	I					I		
10								I	
9			I						
8				I		I			I
7									
6	I	I	I	I	I				
5					I	I	I	I	I
4	II	II	I	II	I	II	I	I	I
3			I						
2					I		I	I	I
1									

Figure 2. Identification of the Critical Regions



* *red asterisks indicate primers used in this study*

* *green asterisks refer to primers used in McLure et al 2005*

NBLST6	1q21.1	Susceptibility to neuroblastoma, 6
PSORS4	1q21	Susceptibility to psoriasis, 4
SLEB14	1q21-23	Susceptibility to systemic lupus erythematosus, 14
HFE2	1q21.1	Hemochromatosis type 2 (juvenile)
MPZ	1q23.3	Myelin protein zero
FCGR3A	1q23	Low affinity IIIa receptor, Fc fragment of IgG (CD16A)

Highlights

- A whole genome approach identifying polymorphic blocks by one of their key features
- The vertebrate genome contains quanta of variation or Polymorphic Frozen Blocks.
- Evidence suggests a new model for primate evolution
- Model based on conservation of polymorphism rather than *de novo* mutation
- Uniqueness is a function of random mixing of haplotypes not random mutation.