European Journal of Human Genetics (2006) 14, 180–189

www.nature.com/ejhg

ARTICLE

Copy number variation in regions flanked (or unflanked) by duplicons among patients with developmental delay and/or congenital malformations; detection of reciprocal and partial Williams-Beuren duplications

Marjolein Kriek*[,1], Stefan J White[1], Karoly Szuhai[2], Jeroen Knijnenburg[2], Gert-Jan B van Ommen[1], Johan T den Dunnen[1] and Martijn H Breuning[1]

[1]*Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands;* [2]*Department of Molecular Cell Biology, Leiden University Medical Center, Leiden, The Netherlands*

Duplicons, that is, DNA sequences with minimum length 10 kb and a high sequence similarity, are known to cause unequal homologous recombination, leading to deletions and the reciprocal duplications. In this study, we designed a Multiplex Amplifiable Probe Hybridisation (MAPH) assay containing 63 exon-specific single-copy sequences from within a selection of the 169 regions flanked by duplicons that were identified, at a first pass, in 2001. Subsequently, we determined the frequency of chromosomal rearrangements among patients with developmental delay (DD) and/or congenital malformations (CM). In addition, we tried to identify new regions involved in DD/CM using the same assay. In 105 patients, six imbalances (5.8%) were detected and verified. Three of these were located in microdeletion-related regions, two alterations were polymorphic duplications and the effect of the last alteration is currently unknown. The same study population was tested for rearrangements in regions with no known duplicons nearby, using a set of probes derived from 58 function-selected genes. The latter screening revealed two alterations. As expected, the alteration frequency per unit of DNA is much higher in regions flanked by duplicons (fraction of the genome tested: 5.2%) compared to regions without known duplicons nearby (fraction of the genome tested: 24.5–90.2%). We were able to detect three novel rearrangements, including the previously undescribed reciprocal duplication of the Williams Beuren critical region, a subduplicon alteration within this region and a duplication on chromosome band 16p13.11. Our results support the hypothesis that regions flanked by duplicons are enriched for copy number variations.
European Journal of Human Genetics (2006) **14,** 180–189. doi:10.1038/sj.ejhg.5201540; published online 14 December 2005

Keywords: duplicons; copy number variation; developmental delay; congenital malformations

Introduction

Many genetic disorders are caused by changes in chromosomal structure. Deletions, duplications, inversions and translocations can all lead to changes in the effective dosage of one or more genes, often with pathological consequences. Large rearrangements affecting at least 5 Mb can be seen cytogenetically, and many disorders have been

*Correspondence: Dr M Kriek, Center for Human and Clinical Genetics, Leiden University Medical Center, Wassenaarseway 72, Leiden, The Netherlands. Tel: +31 71 5276288; Fax: +31 71 5276075;
E-mail: M.kriek@lumc.nl
Received 23 March 2005; revised 17 October 2005; accepted 18 October 2005; published online 14 December 2005

recognised and characterised based solely on microscopic analysis.[1–4]

It was shown in 1992 that the region duplicated in Charcot–Marie–Tooth (CMT) was flanked by highly similar (>98%) sequences.[5] Unequal crossing over between these duplicons leads both to this duplication and the reciprocal deletion, which was later shown to cause hereditary neuropathy with liability to pressure palsies (HNPP).[6] Duplicons, also known as low copy repeats (LCRs), have since been implicated in many other disorders.[7,8] It has been estimated that 5% of the human genome is composed of such LCRs, which can be present both inter- and intrachromosomally.[9,10]

In 2002, Bailey *et al*[11] identified 169 unique regions of at least 10 kb in size, between intrachromosomal duplicons with >95% sequence identity. These data were based on the Human Working draft of August 2001. In all, 24 of these regions were already associated with known genetic disorders. It was hypothesised that these 169 regions are likely to undergo rearrangements more frequently compared to interstitial regions outside the defined regions, due to misaligned recombination between the LCRs, creating microdeletions, microduplications and inversions of the segments involved. To assess this in more detail, we have designed a Multiplex Amplifiable Probe Hybridisation (MAPH) probe set containing 30% of these regions, including those related to microdeletion syndromes. In all, 105 unrelated patients with developmental delay (DD) and/or congenital malformations (CM) were tested using these probes. We compared the performance of this probe set with a set of probes located outside the thus far known duplicons. The second purpose of this study was to identify new regions that are frequently altered in DD patients or patients with CM using the duplicon data of 2002.

The assay using sequences flanked by duplicons resulted in the detection of six duplications, of which three were located in regions related to known disorders. Two alterations were detected by screening regions outside known duplicons. These results show that in our study population the genetic variation within duplicon-flanked regions was three times more common compared to the regions outside the duplicons. Among the rearrangements detected was the postulated, but until now unidentified, reciprocal duplication of the Williams Beuren critical region (WBCR) and a smaller subduplicon alteration within this region.

Materials and methods

Patients

The DNA of 99 DD/CM patients and six individuals with CM only (64 males and 41 females) from the Center of Human and Clinical Genetics Leiden (DNA Diagnostic Laboratory) was analysed. Prior to MAPH analysis, all patients showed a normal karyotype and, where tested, had tested negative for Fragile X syndrome. This study cohort does not include any patient presenting with typical microdeletion characteristics. These had been previously diagnosed by the cytogenetics department.

This study was approved by the Institutional Review Board of the Leiden University Medical Center, conforming to Dutch law. All subjects, or their representatives, gave informed consent for DNA studies.

Multiplex Amplifiable Probe Hybridisation

MAPH was performed as described by White *et al*.[12] Ratios were obtained by dividing the peak height of each probe by the sum of the peak heights of the four nearest probes. The probes with a normalised ratio between 0.75 and 1.25 (log(2) scale −0.42 to +0.32) were considered to be present in two copies. The probes with a ratio outside these thresholds were considered to have a copy number alteration. All samples in which an alteration was found were screened at least in duplicate.

The different probe sets used contained respectively 63 probes from genes flanked by duplicons (see Appendix A) in 51 different regions, including those involved in Smith Magenis (SMS (MIM 182290)), William Beuren (WBS (MIM 194050)), DiGeorge (DGS (MIM 188400)), Cat eye (CES (MIM 115470)), Prader Willi (PWS (MIM 176270)), Angelman syndrome (AS (MIM 105830)) and 58 probes containing function-selected genes outside the duplicons (Appendix B).

Multiplex Ligation-dependent Probe Amplification

A modified protocol of multiplex ligation-dependent probe amplification (MLPA)[13] was performed as described by White *et al*.[14] In the current study, MLPA was performed to verify alterations obtained by MAPH analysis. The data analysis is identical with that applied for MAPH analysis. The MLPA probes used were derived from the sequences of *RAI1* (GeneID: 10743), *DRG2* (GeneID: 1819), *COPS3* (GeneID: 8533), *ELN* (GeneID: 2006), *CYLN2* (GeneID: 7461), *FKBP6* (GeneID: 8468), *TBL2* (GeneID: 26608), *FZD9* (GeneID: 8326), *GTF2IRD1* (GeneID: 84163), *GTF2I* (GeneID: 2969), *HIP1* (GeneID:3092), *AUTS2* (GeneID:26053), *CALN1* (GeneID: 83698), *NUDE1* (GeneID: 54820), *PYRR1*, defender against cell death 1 (*DAD1*) gene (GeneID: 1603) and the diacylglycerol kinase iota (*DGKI*) gene (GeneID: 9162).

Fluorescence *In Situ* Hybridisation

The FISH experiments were performed following Standard Operating Procedures.[15] An FITC-labeled FISH clone LSI-ELN (Vysis) was used for the Williams critical Region. BAC clones RP11-14N9, RP11-M13, RP11-489O1 and RP11-72I8 were used to determine the extent of the rearrangement on chromosome band 16p13.3.

Array comparative genomic hybridisation

The array comparative genomic hybridisation (array-CGH) procedures were performed as described in Knijnenburg *et al*[16] using larger genomic insert clones retrieved from the Sanger Center (UK) (1 MB clone set). *In silico* data at the http://www.ensemble.org were used to determine the size of the duplications.

Results

Considering that duplicon-flanked regions might be preferentially involved in copy number variation, we based our MAPH probe set to detect new regions involved in DD/CM on a gene-enriched selection from the 169 regions published by Bailey *et al*.[11]

The MAPH probes were designed based on autosomal exon-specific single-copy sequence. Regions lacking known genes and/or single-copy sequence (62/169 or 37% of the defined regions) were excluded. Before the actual screening, the probe sets were validated using DNA samples derived from 50 anonymous healthy controls. Among those, we detected a pancreatic polypeptide receptor 1 (*PPYR1*) gene duplication that was verified using MLPA analysis. Probes showing inconsistent copy number variation within an individual (duplicate testing) were excluded ($n=9$). The validated probe sets, targeting 63 unique sequences in 51 different regions (see Appendix A), were tested among a total of 105 unrelated patients (64 males, 41 females), including 99 developmentally delayed (DD) patients (25 mild DD; 74 severe DD) and six individuals with CM.

Screening these 105 patients revealed six imbalances (5.8%), all duplications (Table 1). All rearrangements were verified using MLPA, array-CGH or FISH. Three of the rearrangements were located in areas known to be involved in microdeletion syndromes, including two duplications within the WBCR on chromosome band 7q11.23 (see case reports), and a *de novo* duplication of the Smith Magenis Critical Region (SMCR) on chromosome band 17p11.2. The two 7q11.23 duplications, detected in two unrelated patients, differed in length, as one was found using four MAPH probes (containing sequences derived from the *CYLN-2, ELN, FKBP6* and *TBL2* genes) and the other with only one of these, the *FKBP6* gene (Figure 1). Additional array-CGH analysis did not detect this alteration. The exact size of the duplication is difficult to define as the BACs flanking this region (RP11-450O3, RP4-771P4) partly colocalise with segmental duplicons in this region. Additional MLPA was performed using sequences of the *GTF2I* and *GTF2IRD1* genes within the WBCR and *HIP1*, *CALN1* and *AUTS2* genes localised just outside the telomeric and centromeric sides of the segmental duplicon, respectively. This assay revealed that this duplication is the reciprocal duplication of the deletion causing Williams–Beuren syndrome.

To fine map the other duplications (case 2), additional MLPA probes were designed. Exon 4 and exon 8 (the last exon) of the *FKBP6* gene were shown to be duplicated. We were unable to test the first three exons of this gene, as they contain large repetitive sequences. The probe derived from the adjacent *FZD9* gene showed no alteration. Testing the parents of the patients showed that in each case the duplication was present in one of the parents (data not shown). There appeared to be no parent of origin effect, as the large alteration was found in the patient's father, and the small alteration in the mother of the other patient.

The duplication of the SMCR (case 3) was detected using three probes corresponding to the *RAI1, DRG2* and *COPS3* gene. Array-CGH testing was performed to determine the length of the duplication on chromosome 17 (Table 1). This analysis excluded a duplication of chromosome band 17p12, which causes CMT disease (Figure 2).

Chromosome 16 contains many repeats, limiting the application of additional FISH analysis. Thus, it was not possible to determine the precise breakpoints of the imbalance in case 4, a *de novo* duplication of the *NUDE1*

Table 1 Alterations in regions flanked by duplicons

Case	*Alteration*	*Chrom. Band*	*Gene(s) involved*	*Size (Mb)*	*de novo*	*Confirmed by*
Case 1	Duplication	7q11.23	CYLN2, ELN, FZD9,FKBP6, TBL2	1.4–1.7	No, present in father	MLPA/FISH
Case 2	Duplication	7q11.23	FKBP6	0.3–0.4	No, present in mother	MLPA
Case 3	Duplication	17p11.2	RAI1, DRG2, COPS3	min. 3.5[a]	Yes	MLPA/FISH/array-CGH
Case 4	Duplication	16p13.11	NUDE1, MYH11	0.8–2.4	Yes	MLPA/FISH/array-CGH
Case 5	Duplication	10q11.22	PPYR1	0.5–2.3	No, present in father	MLPA/array-CGH
Case 6	Duplication	10q11.22	PPYR1	max. 1.4	Unknown[b]	MLPA/array-CGH

Summary of results obtained by screening 105 DD/CM patients using 51 unique regions flanked by duplicons. The sizes of the different alterations were determined based on results of both MAPH/MLPA and array-CGH.
[a]As the regions near the centromere of chromosome 17 are not covered by array-CGH, the centromeric breakpoint of this duplication remains unknown.
[b]The mother of case 6 did not carry the duplication. The father was not available for testing.

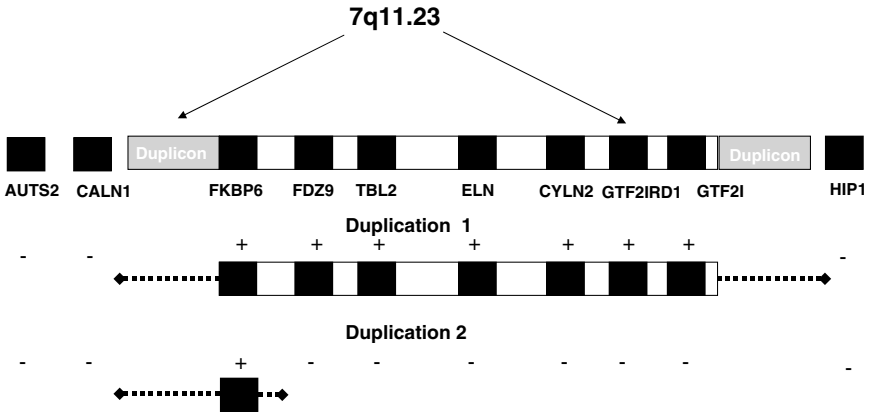

Figure 1 The duplications within 7q11.23 (WBCR). The figure shows the length of the two duplications in the WBCR, detected in unrelated patients. Duplication 1 encompasses the whole critical area flanked by two large duplicons, whereas the other duplication involves only (a part of) the *FKBP6* gene. The diamonds represent the maximum size of both duplications. The *AUTS2, CALN1* and *HIP1* genes localised just outside the duplicons were not altered.

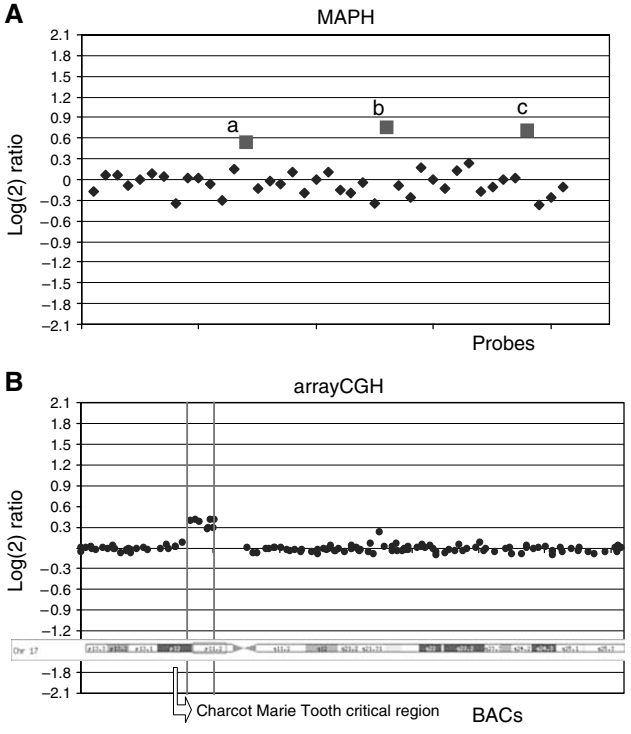

Figure 2 Results obtained in case 3. Results of the MAPH and array-CGH analysis revealing a duplication of the SMCR. (**A**) Log(2) ratio of MAPH probes showing a duplication of (a) the *RAI1* gene, (b) the *DRG2* gene and (c) the *COPS3* gene. The remaining probes contained sequences localised on different chromosomes. The probes with a normalised ratio between −0.42 and +0.32 (log(2) scale) were considered to be present in two copies. The probes are ordered by probe length, not on their position on the genome. (**B**) Array-CGH testing showed that chromosome band 17p12 is not duplicated, excluding CMT syndrome (white arrow). The BACs showing amplification included RP11–219A15, RP11–524F11, RP11–189D22, RP1–162E17, CTB–1187M2, RP11–78O7, RP5–836L9 and RP11–121A13. The distal breakpoint matches the common deletion breakpoint of SMS.[18] The proximal breakpoint is unknown, as the region near the centromere is not covered by BACs.

gene on the short arm of chromosome 16p13.11. Two BACs (RP11-489O1, CTD-2504F3) overlapping the *NUDE1* region were found amplified using array-CGH, indicating that the size of the duplication is between 0.8 and 2.4 Mb. We note that the dosage of the *MYH11* gene (Locus Link: 4629) must also be doubled as this gene is transcribed from the reverse strand of the *NUDE1* gene.

In two unrelated patients (cases 5 and 6), a duplication of a probe within the first exon of the *PPYR1* gene on chromosome 10 was identified and subsequently verified using MLPA. Using array-CGH analysis, a nonoverlapping BAC (RP11-292F22) localised 0.5 Mb telomeric from the *PPYR1* gene showed a duplication in only one of the patients, indicating a difference in the size of the regions duplicated. We were able to test both parents of the patient with the largest rearrangement (case 5); the father carried the same duplication. The mother of the other patient did not show the duplication, the father was not available for testing.

To determine whether the number of alterations obtained is significantly higher compared to copy number changes of regions outside the duplicons described in 2001, we have tested the same study population for genomic variation in a set of probes from regions not known to be flanked by duplicons. These probes were targeting function-selected genes, such as genes involved in transcription, neuronal and brain maturity, with a potential function in mental development (Appendix B). This MAPH analysis comprised 58 validated probes (Appendix B) and resulted in the detection of two genetic imbalances (1.9%), including a duplication of the *DGKi* gene on chromosome band 7q33 and a deletion of the *DAD1* gene on chromosome band 14q11. Both alterations were verified by MLPA analysis. We were not able to test the parents of these patients. Despite their predicted function, these genes have not previously been causally linked to DD.

Case reports

Case 1 This male patient was born after an uneventful pregnancy. In the perinatal period, he was diagnosed with trigonocephalic synostosis of the metopic ridge. At the age of 1 year, he was examined by a clinical geneticist. He did not show any DD nor obvious dysmorphic features. Except for a mild aberrant shape of his skull (status after reconstruction), no CM were present.

The family history of this patient included, in the father with a complete cutaneous III–IV syndactyly of the hand, a II–III syndactyly of the feet, and a carcinoma *in situ* of the testis that was diagnosed after infertility screening. The family members of both the father's mother and father's father showed syndactyly. Additional MAPH analysis showed a duplication of the WBCR present in the patient as well as in the father. The parents of the patient's father did not carry the duplication. The parenthood of the father and his parents was proven using marker studies.

Case 2 In addition to synostosis of both the sutura lamboidea and the sutura coronalis, this 4-year-old male patient with a normal mental development showed facial asymmetry, a severe heart malformation including two ventricular septum defects and a (sub)valvular pulmonal stenosis and a finger-like thumb. Except for craniosynostosis, these features are related to hemifacial microsomia.

The family history does not include individuals with dysmorphic features nor CM. Additional investigation showed a normal karyotype. MAPH analysis showed a duplication of a part of the *FKBP6* gene that was also present in the unaffected mother and the unaffected maternal grandmother.

Discussion

In this study, we have assessed the frequency of chromosomal rearrangements in DD and/or CM patients. The fraction of the genome that was localised between the defined duplicons (as of 2001) and tested by at least one MAPH probe was 5.2% (see Appendix A). Within these regions, six alterations were detected. The fraction of the genome that was flanked by duplicons and not tested in this study was 4.6%, indicating that the majority of the genome fraction flanked by duplicons has been tested in this study. The total fraction of the genome that was flanked by duplicons identified at a first pass in 2001 is thus 9.8%. This percentage corresponds closely with the ~328 Mb of sequence calculated by Bailey *et al*.

The fraction of the genome unflanked by duplicons (defined in 2001) is 90.2%. However, we have only tested 58 sequences (probes) localised outside the duplicons. We would argue that this number is not representative for 90.2% of the genome. Based on the calculation shown in Appendix B, the fraction of the non-duplicon regions tested was at least 24.5%. The real percentage tested is higher, as sequences located at the chromosome ends could not be included. In short, the fraction of the genome localised outside the duplicons and tested ranges between 24.5 and 90.2%. Two alterations were found within these regions. While the sample sizes are small, the aberration frequency per unit (=percentage of the total genome) of DNA in regions flanked by duplicons was higher compared to the regions outside the duplicons, indicating that the regions between the duplicons are indeed enriched for dosage alterations. This supports the hypothesis of Bailey *et al*. that the regions within duplicons are more likely to undergo genomic alterations.

Retrospectively, we have checked all 58 genes localised outside the duplicons, as identified in 2001, using the most recent assembly of the Human Working Draft (May 2004). It appeared that 76% of these regions were still unflanked by intrachromosomal duplicons, including the regions containing *DGKi* and *DAD1* genes.

Several factors will lead to an underestimation of the true number of alterations occurring between duplicons, and some of these may also explain why we did not find any deletions. First, the regions lacking single-copy sequences were excluded in this study. It is reasonable to assume that these regions are more likely to undergo rearrangements based on their repetitive sequence content. These were not included, as the MAPH assay was based on copy number alteration of single-copy sequences.

Second, haplo-insufficiency of certain genes might not be compatible with life, or they may give a deleterious phenotype other than DD/CM. These alterations will not be detected in our study. This holds equally for the function-selected genes. Brewer *et al*[17] defined several regions that have never been involved in any deletion and those were thought to be potentially haplo-lethal. Of the 57 'Bailey' regions tested, 10 were located within these possible haplo-lethal regions. These regions need to be tested by higher resolution methods, as the analysis of Brewer *et al* was based on karyotypic abnormalities. Third, a substantial proportion of DD/CM could originate from genetic aberrations other than nonallelic homologous recombination. For example, point mutations will not be detected using MAPH.

Fourth, the number of samples tested is rather small and the set of probes outside the duplicons is not random. In addition, the study cohort is already biased against rearrangements between duplicons, as any cases presenting with typical microdeletion syndrome-related features had already been diagnosed using cytogenetics tools.

Finally, it is possible that a part of the duplicons defined by Bailey *et al* require additional conditions before the obligate 'repetitive breakpoints events' will occur, resulting in copy number changes. These additional conditions could include a minimum length of 100% homology required for recombination, AT-rich sequences present on both sites of a recombination hotspots,[18] or enrichment of

Alu repeats within duplicons.[19] Further analysis needs to be performed to determine whether these conditions are present in the 'Bailey'-defined duplicons.

A more clinical question concerns whether the imbalances found are disease-causing changes or benign polymorphisms. Alterations due to misaligned nonallelic homologous recombination should result in a deletion and a reciprocal duplication. In the majority of reciprocal deletion/duplication disorders, deletions were discovered before the duplication of the regions due to the fact that the techniques applied (usually FISH) were more amenable for deletion detection. To date, several duplications in regions involved in microdeletion syndromes have been identified in addition to the known deletions.[20–23] The phenotype corresponding to the duplication is often milder than that related to the deletion. However, the copy number changes can also be associated with polymorphic variation.[24]

Due to the presence of >320 kb repeat structure on both sides of the Williams syndrome critical region, the existence of a reciprocal duplication of the Williams critical region was predicted,[25,26] however, it has not been reported before. The patient with the reciprocal duplication of the Williams critical region was diagnosed with craniosynostosis and mild DD. The patient with the smaller duplication showed, in addition to craniosynostosis, multiple CM; however, his psychological development was normal. As the *FKBP6* gene is the only gene in common and this gene is restricted to the male germ cells, it is reasonable to assume that the clinical overlap (craniosynostosis) is coincidental.

The clinical consequences of a duplication within the WBCR are currently unknown. The fact that the imbalance is present in unaffected family members does not automatically mean that this is not pathological. Incomplete penetrance or multifactorial influences might cause variability of the phenotype.

It seems reasonable to assume that the *de novo* 17p11.2 duplication is responsible for the clinical features of case 3, as it is known that a duplication of the SMS critical region is associated with clinical features resembling those observed in our patient.[23,27]

The *de novo* duplication of 16p13.11 was seen in a boy with mild DD and learning disability. Since the father had similar learning problems, the significance of the duplication is questionable and this awaits confirmation from other patients. We note, however, that *NUDE1* participates in a pathway that influences the neuronal migration during development of the central nervous system,[28] which makes it an interesting candidate gene in this region.

Sebat *et al*[29] reported the screening of a total of 20 healthy individuals using the representational oligonucleotide microarray analysis (ROMA) technique. They found 76 unique large-scale copy number polymorphisms. Among those, five probes on chromosome band 10q11.2 encompassing the full length of the *PPYR1* gene were duplicated in one individual. This finding is in agreement with our finding of no less than four copy number changes in this gene, as it was altered in two unrelated patients (cases 5 and 6), one of their parents, as well as in a healthy control sample. In a subsequent study regarding genomic copy number differences in healthy individuals, 255 loci showing large-scale copy number variation (LCVs) were detected using array-CGH analysis.[30] The only probe that overlapped one of the 255 suspected polymorphic clones contained a *PPYR1* gene sequence. This clone (AL390716.27) was amplified in six individuals. Combining these findings in retrospect, it is possible that *PPYR1* undergoes nonpathological or incompletely penetrant copy number variation. Two of the function-selected genes were localised within the suspected polymorphic clones (*RYR3* within clone ACO11938.4; *ERN1* within clone RP11-89H15). The probes derived from both genes were not altered in our study population. This may well be due to our modest sample size, since most copy number variations detected by Iafrate *et al* were present in only one or two (healthy) individuals. This also holds true for the clones overlapping *RYR3* and *ERN1*. In addition, a duplication seen with a single BAC clone might not encompass the entire clone length.

Recently, Sharp *et al*[31] also found a difference with regard to duplicons-flanked regions and copy number variation, in agreement with our findings. In addition, 130 potential copy number variation hotspots flanked by duplicons were tested for rearrangements among 47 healthy individuals using a segmental duplicon BAC microarray. A total of 119 regions showed copy number alteration comprising 141 genes, including the *P25*, *P29* and *ADRBK2* genes, also present in our study. In all, 79 of the 130 copy number variation hotspots showed no alteration among this study population. It was suggested that these latter hotspots are excellent candidate regions to be associated with genetic disorders. Our study covers a fraction of these 'hotspots', which have thus been subjected to a first test for copy number alteration in relation to DD or CM. Using MAPH, we were able to identify three previously undescribed rearrangements, two duplications within WBCR and one duplication of chromosome region 16p13.11, of which the clinical relevance is uncertain at this moment. It will indeed be worthwhile to include these regions in further testing.

Acknowledgements

We thank Hans Dauwerse, Kerstin Hansson, Jeroen Nijhuis for the FISH analysis, Yvonne Hilhorst for providing clinical information, Peter de Knijff for parental marker analysis. MK is funded by Zon-Mw (AGIKO fellowship 940-37-032), SW is funded by ZonMw (nr 912-04-047).

Electronic-Database Information

- *Leiden Muscular Dystrophy Pages, http://www.dmd.nl/DMD_MAPH.html*

- *Prophet (http://www.basic.nwu.edu/biotools/prophet.html)*
- *UCSC Genome Browser (http://genome.ucsc.edu/)*
- *Ensembl (http://www.ensembl.org)*
- *Online Mendelian Inheritance in Man (OMIM), http://www. ncbi. nlm.nih.gov/Omim*
- *Genome Variation Database, http://projects.tcag.ca/variation/.*

Note added in proof

While this work was under review, another patient was described (Severe expressive-language delay related to duplication of the Williams-Beuren locus, MJ Somerville et al. N Engl J Med 2005; ***353****:1694–1701, October 20, 2005) with a duplication of the WBS region. We have assessed the phenotype of our patient in the light of the reported clinical features (language deficiency but good spatial abilities). Considering the age of our patient, we could not assess the spatial abilities, but our patient did present with (moderate) language disability.*

References

1 Leao JC, Bargman GJ, Neu RL, Kajii T, Gardner LI: New syndrome associated with partial deletion of short arms of chromosome No. 4. Clinical manifestations of hypospadias, beaked nose, abnormal iris, hemangioma of forehead, seizures, and other anomalies. *J Am Med Assoc* 1967; **202**: 434–437.
2 Alfi O, Donnell GN, Crandall BF, Derencsenyi A, Menon R: Deletion of the short arm of chromosome no.9 (46,9p-): a new deletion syndrome. *Ann Genet* 1973; **16**: 17–22.
3 Schinzel A, Auf der MP, Moser H: Partial deletion of long arm of chromosome 11[del(11)(q23)]: Jacobsen syndrome. Two new cases and review of the clinical findings. *J Med Genet* 1977; **14**: 438–444.
4 Greenberg F, Crowder WE, Paschall V, Colon-Linares J, Lubianski B, Ledbetter DH: Familial DiGeorge syndrome and associated partial monosomy of chromosome 22. *Hum Genet* 1984; **65**: 317–319.
5 Lupski JR, Wise CA, Kuwano A et al: Gene dosage is a mechanism for Charcot–Marie–Tooth disease type 1A. *Nat Genet* 1992; **1**: 29–33.
6 Chance PF, Abbas N, Lensch MW et al: Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. *Hum Mol Genet* 1994; **3**: 223–228.
7 Emanuel BS, Shaikh TH: Segmental duplications: an 'expanding' role in genomic instability and disease. *Nat Rev Genet* 2001; **2**: 791–800.
8 Stankiewicz P, Lupski JR: Genome architecture, rearrangements and genomic disorders. *Trends Genet* 2002; **18**: 74–82.
9 Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE: Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res* 2001; **11**: 1005–1017.
10 Eichler EE: Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet* 2001; **17**: 661–669.
11 Bailey JA, Gu Z, Clark RA et al: Recent segmental duplications in the human genome. *Science* 2002; **297**: 1003–1007.
12 White S, Kalf M, Liu Q et al: Comprehensive detection of genomic duplications and deletions in the DMD gene, by use of multiplex amplifiable probe hybridization. *Am J Hum Genet* 2002; **71**: 365–374.
13 Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G: Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res* 2002; **30**: e57.
14 White SJ, Vink GR, Kriek M et al: Two-color multiplex ligation-dependent probe amplification: detecting genomic rearrangements in hereditary multiple exostoses. *Hum Mutat* 2004; **24**: 86–92.
15 Dauwerse JG, Jumelet EA, Wessels JW et al: Extensive cross-homology between the long and short arm of chromosome 16 may explain leukemic inversions and translocations. *Blood* 1992; **79**: 1299–1304.
16 Knijnenburg J, Szuhai K, Giltay J et al: Insights from genomic microarrays into structural chromosome rearrangements. *Am J Med Genet A* 2005; **132**: 36–40.
17 Brewer C, Holloway S, Zawalnyski P, Schinzel A, Fitz Patrick D: A chromosomal duplication map of malformations: regions of suspected haplo- and triplolethality – and tolerance of segmental aneuploidy – in humans. *Am J Hum Genet* 1999; **64**: 1702–1708.
18 Shaw CJ, Withers MA, Lupski JR: Uncommon deletions of the Smith–Magenis syndrome region can be recurrent when alternate low-copy repeats act as homologous recombination substrates. *Am J Hum Genet* 2004; **75**: 75–81.
19 Bailey JA, Liu G, Eichler EE: An Alu transposition model for the origin and expansion of human segmental duplications. *Am J Hum Genet* 2003; **73**: 823–834.
20 Lupski JR: Charcot–Marie–Tooth disease: lessons in genetic mechanisms. *Mol Med* 1998; **4**: 3–11.
21 Shaw-Smith C, Redon R, Rickman L et al: Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. *J Med Genet* 2004; **41**: 241–248.
22 Ensenauer RE, Adeyinka A, Flynn HC et al: Microduplication 22q11.2, an emerging syndrome: clinical, cytogenetic, and molecular analysis of thirteen patients. *Am J Hum Genet* 2003; **73**: 1027–1040.
23 Potocki L, Chen KS, Park SS et al: Molecular mechanism for duplication 17p11.2 – the homologous recombination reciprocal of the Smith–Magenis microdeletion. *Nat Genet* 2000; **24**: 84–87.
24 Stankiewicz P, Lupski JR: Molecular-evolutionary mechanisms for genomic disorders. *Curr Opin Genet Dev* 2002; **12**: 312–319.
25 Bayes M, Magano LF, Rivera N, Flores R, Perez Jurado LA: Mutational mechanisms of Williams–Beuren syndrome deletions. *Am J Hum Genet* 2003; **73**: 131–151.
26 Peoples R, Franke Y, Wang Y et al: A physical map, including a BAC/PAC clone contig, of the Williams–Beuren syndrome–deletion region at 7q11.23. *Am J Hum Genet* 2000; **66**: 47–68.
27 Shaw CJ, Bi W, Lupski JR: Genetic proof of unequal meiotic crossovers in reciprocal deletion and duplication of 17p11.2. *Am J Hum Genet* 2002; **71**: 1072–1081.
28 Yan X, Li F, Liang Y et al: Human Nudel and NudE as regulators of cytoplasmic dynein in poleward protein transport along the mitotic spindle. *Mol Cell Biol* 2003; **23**: 1239–1250.
29 Sebat J, Lakshmi B, Troge J et al: Large-scale copy number polymorphism in the human genome. *Science* 2004; **305**: 525–528.
30 Iafrate AJ, Feuk L, Rivera MN et al: Detection of large-scale variation in the human genome. *Nat Genet* 2004; **36**: 949–951.
31 Sharp AJ, Locke DP, McGrath SD et al: Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 2005; **77**: 78–88.

Appendix A.

See Table A1.

Appendix B.

See Table B1.

Table A1 An overview of 63 genes tested using MAPH analysis among 105 DD/CM patients

Gene	*Chrom. band*	*Description*	*GeneID*	*Location on chromosome*	*Regions flanked by duplicons*	*Distance between duplicon*
P29	1p35.3	GCIP-interacting protein p29	25949	28755625	28834059–28932575	98 516
PRKAB2	1q21.1	Protein kinase, AMP-activated, beta 2	5565	173584889	171036140–177249592	62 13 452
CAPN2	1q42.11	Calpain 2, large subunit	824	259601508	258169103–259907561	1 738 458
FLJ2204	2q13	Hypothetical protein FLJ22004		117458997	117178230–117635198	456 968
NPHP1	2q13	Nephrocystin	4867	114316563	109925083–116080082	6 154 999
UMPS	3q21.2	Uridine monophosphate synthetase (orotate)	7372	141469386	140479577–143078865	2 599 288
GLUC	4p15.2	Cytosolic beta-glucosidase	2629	25264898	24981138–25427505	446 367
EVC	4p16.2	Ellis van Creveld syndrome protein	2121	6019414	4420622–10868121	6 447 499
P25	5p15.33	Brain-specific protein p25 alpha	11076	1251194	1180464–1306142	125 678
RANBP17	5q35.1	RAN binding protein 17	64901	188–189 Mb	189063686–189213481	149 795
MLN	6p21.31	Motilin	4295	37345065	34908829–40172081	5 263 252
DDC	7p12.2	Dopa decarboxylase (aromatic L-amino acid)	1644	55065653	49777336–61172562	11 395 226
GSBS	7p14.3	G-substrate	10842	34265228	31828578–37665179	5 836 601
JTV1	7p22.1	Multisynthetase complex auxiliary component	7965	6767988	6652376–7723723	1 071 347
TPST1	7q11.21	Tyrosylprotein sulphotransferase 1	8460	68961571	68899125–69000004	100 879
FKBP6[a]	7q11.23	FK506-binding protein 6	8468	70913203	70865853–71592416	726 563
TBL2[a]	7q11.23	Transducin (beta)-like 2	26608	*	70865853–71592416	see above
ELN[a]	7q11.23	Elastin	2006	*	70865853–71592416	see above
CYLN2[a]	7q11.23	Cytoplasmic linker 2	7461	*	70865853–71592416	see above
ARHGEF5	7q35	Rho guanine nucleotide exchange factor 5	7984	156140962	155961558–156151892	190 334
CENTG3	7q36.1	MRIP-1 protein	116988	163305750	161981314–166427628	4 446 314
di-RAS2	9q22.2	GTP-binding RAS-like 2	54769	99459475	99072373–99615625	543 252
PTCH	9q22.32	Patched (*Drosophila*) homolog	5727	107463432	105649233–108546485	2 897 252
FANCC	9q22.33	Fanconi anaemia, complementation group C	2176	107816739	105649233–108546485	See above
RSU1	10p13	ras suppressor protein 1	6251	17763985	17343558–17953656	610 098
KIAA0187	10q11.21	KIAA0187 gene product	9790	45421390	45331297–47906414	2 575 117
SDF1	10q11.21	Stromal cell-derived factor 1	6387	47064383	45331297–47906414	See above
PPYR1	10q11.22	Pancreatic polypeptide receptor 1	5540	49145372	49021238–54773984	5 752 746
SGPL1	10q22.1	Sphingosine-1-phosphate lyase 1	8879	77278049	75454903–80191770	4 736 867
TACR2	10q22.1	Tachykinin receptor 2	6865	75719214	75454903–80191770	See above
PAPSS2	10q23.31	3′ phosphoadenosine 5′ phosphosulphate	9060	88057489	86760491–88848332	2 087 841
FLJ22794	11q12.1	Hypothetical protein FLJ22794	63901	65208275	65086253–65169286	83 033
CD5	11q12.2	CD5 antigen (p56–62)	921	67565428	67348654–67606043	257 389
FADD	11q13.3	Fas (TNFRSF6)-associated via death domain	8772	79700968	77157096–82393752	5 236 656
ICEBERG	11q22.3	ICEBERG caspase-1 inhibitor	59082	120243026	120179983–120328323	148 340
HNT	11q25	Neurotrimin precursor	50863	151230535	151288128–151483616	195 488
CLECSF12	12p13.2	C-type lectin domain family c7, member Ca dep.	64581	10905383	10753561–11533368	779 807
CNTN1	12q12	Contactin 1	1272	45800931	45700335–47233112	1 532 777
DKFZp434B0417	12q12	Hypothetical protein DKFZp434B0417		46680635	45700335–47233112	see above
TMEM5	12q14.2	Transmembrane protein 5	10329	72955978	72479865–73099895	620 030
CKAP2	13q14.3	Cytoskeleton-associated protein 2	26586	52929838	52126921–53081328	954 407
NDN[b]	15q11.2	Necdin	4692	19787505*	17304292–19469943	2 165 651
UBE3A[c]	15q12	Ubiquitin protein ligase E3A	7337	21515963*	20279911–20507618	227 707
LTK	15q15.1	Tyrosine kinase	4058	38501171	35955283–41645045	5 689 762
NMB	15q25.2	Neuromedin B	4828	88362943	85255238–88736771	3 481 533
NADRIN	16p12.1	Neuronal protein	55114	29985255	14335149–35125392	20 790 243
NUDE1	16p13.11	Lis-1 interacting protein	54820	18763116	14335149–35125392	See above
TAT	16q22.2	Tyrosine aminotransferase	6898	85891517	82945639–89466425	6 520 786
CFDP1	16q23.1	Craniofacial development protein 1	10428	90635426	90575990–90735398	159 408
DRG2[d]	17p11.2	Developmentally regulated GTP binding protein 2	1819	19787405	15371266–27948279	12 577 013
COPS3[d]	17p11.2	*Homo sapiens* COP9 complex subunit 3	8533	19038181	15371266–27948279	See above

Table A1 (*Continued*)

Gene	*Chrom. band*	*Description*	*GeneID*	*Location on chromosome*	*Regions flanked by duplicons*	*Distance between duplicon*
RAI1[d]	17p11.2	Retinoic acid induced 1	10743	19492572	15371266–27948279	See above
NF1	17q11.2	Neurofibromin	4763	32548362	31949051–33721569	1 772 518
ACACA	17q12	Acetyl-coenzyme A carboxylase alpha	31	38913111	37945776–39868543	1 922 767
ASPA	17p13.2	Aspartoacylase	443	3267932	3120079–3546982	426 903
CLTC	17q23.2	Clathrin heavy chain	1213	65270461	65066121–65736364	670 243
TBX2	17q23.2	T-box 2	6909	67106821	65739747–68308666	2 568 919
IMPA2	18p11.21	Inositol(myo)-1(or 4)-monophosphatase 2	3613	13146682	12188020–13392618	1 204 598
LIPG	18q21.1	Endothelial lipase precursor	9388	54173714	54104840–54449609	344 769
FLJ14686	19q13.12	Zinc-finger protein 382	84911	44915972	44764587–46070350	1 305 763
NOSIP	19q13.33	Nitric oxide synthase interacting protein	51070	61526455	59278949–62006526	2 727 577
SPIB	19q13.33	Spi-B transcription factor (Spi-1/PU.1 related)	6689	62489290	62019412–62726350	706 938
ECR2[e]	22q11.1	Cat eye syndrome chromosome region, candidate	27443	14900358	13950072–21770926	7 820 854
DGCR2[f]	22q11.2	DiGeorge syndrome critical region gene 2	9993	15882238	13950072–21770926	See above
ADRBK2	22q12.1	Beta adrenergic receptor kinase 2	157	22657045	13950072–21770926	See above

The probes were designed using exon-specific single-copy sequences located in regions defined at a first pass by Bailey *et al*.[11] The localisation of the sequences is based on the Human Working draft of August 2001, as the duplicon data of Bailey is based on this information. Some of the probes tested were localised within the regions related to microdeletion syndromes: [a]Williams syndrome. [b]Prader Willi. [c]Angelman syndrome. [d]Smith Magenis syndrome. [e]Cat eye syndrome. [f]22q11del/dup syndrome. *These genes are now known to be located in the WBCR; however, in the Human Working Draft of August 2001, these genes were located outside the Williams-related duplicons.
The sum of all basepairs that are localised between two homologous intrachromosomal duplicons and tested in this study is 155 556 588 bp. This resembles 5.2% of the total human genome. The sum of all basepairs localised between duplicons and not tested in this study is 4.6% (calculation not shown). The total percentage of the genome flanked by duplicons identified at a first pass in 2001 is 9.8%.

Table B1 An overview of 58 probes containing function-selected genes localised outside the duplicons

Gene	*Chrom. band*	*Description*	*GeneID*	*Location on chromosome*	*Interval regions outside duplicons*	*Distance between nearest duplicons*
MATN3	2p24.1	Matrilin 3	4148	20824361	Nearest 92015946	
FACL3	2q35	Acyl-CoA synthetase long-chain family	5147	233250166	137736981–242791383	105 054 402
PDE6D	2q37.1	Phosphodiesterase	7182	242633293	137736981–242791383	See above
NR2C2	3p25.1	Nuclear receptor subfamily	8087	20380273	Nearest 61700285	—
FXR1	3q26.33	Fragile X mental retardation	6750	206812624	204266505–223161158	18 894 653
SST	3q27.3	Somatostatin	10934	214072357	204266505–223161158	See above
MORF4	4q34.1	Mortality factor 4	27295	190237096	158224519–207089932	48 865 413
ALP	4q35.1	Actinin-assoc. protein	10409	203065626	158224519–207089932	See above
BASP1	5p15.1	Brain abundant, membrane-attached signal protein	2554	20241923	1306142–20506502	19 200 360
GABRA1	5q34	GABA receptor	3720	179091963	122206614–189063686	66 857 072
JMJ	6p22.3	Jumonji	9113	17824722	Nearest 28850598	—
LATS1	6q25.1	Tumour suppressor	4697	169320238	104572925–191797029	87 224 104
NDUFA4	7p21.3	NADH dehydrogenase	9162	11791714	7723723–31828578	24 104 855
DGK1	7q33	Diacylglycerol kinase oita	6456	148071336	138492661–155481235	16 988 574
SH3GL2	9p22.2	SH3-domain GRB2-like 2	80380	19168227	Nearest 37513397	—
PDL2	9p24.1	Programmed cell death 1 ligand 2	7099	5822952	Nearest 37513397	—
TLR4	9q33.1	Toll-like receptor 4	7248	130092974	Nearest 108546485	—
TSC1	9q34.13	Tuberous sclerosis 1	6812	146741930	Nearest 108546485	—
STXBP1	9q34.13	Syntaxin-binding protein	64376	141343041	Nearest 108546485	—
PEGASUS	10q26.12	Zinc-finger protein, subfamily 1A, 5	372	135434807	Nearest 86760491	—

Table B1 (*Continued*)

Gene	*Chrom. band*	*Description*	*GeneID*	*Location on chromosome*	*Interval regions outside duplicons*	*Distance between nearest duplicons*
HCCA2	11p15.5	YY1 associated protein	55249	649519	Nearest 3676771	—
ARCN1	11q23.3	Archain 1	6734	134627697	120328323–151288128	30 959 805
SRPR	11q24.2	Signal recognition particle receptor	93661	144635923	120328323–151288128	See above
CAPPA3	12p12.3	Actin-assoc. protein	10959	20564435	12446880–38117363	25 670 483
RNP24	12q24.31	Coated vesicle membrane protein	7223	143196772	Nearest 73099895	—
TRPC4	13q14.11	Transient receptor potential cation channel	2073	36935467	22748066–52126921	29 378 855
ERCC5	13q33.1	Excision repair cross-complementing rodent repair deficiency	1948	106470199	Nearest 64718332	—
EFNB2	13q34	Ephrin-B2	1603	110956506	Nearest 64718332	—
DAD1	14q11.2	Defender against cell death 1	801	19506399	16665813–20896466	4 230 653
CALM1	14q32.11	Calmodulin 1	6263	89723561	Nearest 20979168	—
RYR3	15q14	Ryanodine receptor	27023	29382527	28243975–30470202	2 226 227
FOXB1	15q22.2	Forkhead box 1	3073	58697843	42719271–72062958	29 343 687
HEXA	15q23	Hexosaminidase A	3419	71660617	42719271–72062958	See above
IDH3A	15q24.3	Isocitrate dehydrogenase	98994828	79869724	75343377–85255238	9 911 861
SV2B	15q26.1	Synaptic vesicle protein	53739899	96312975	88736771–105696099	16 959 328
PMM2	16p13.2	Phosphomannomutase	64775373	10760222	4249026–14335149	10 086 123
SIAH1	16q12.1	Cell cycle control	43136477	57195286	38606337–82945639	44 339 302
MMP2	16q12.2	Metalloproteinase; collagen cleavage	70844313	65631286	38606337–82945639	See above
TK2	16q22.1	Mitochondrial thymidine kinase	40947084	79020967	38606337–82945639	See above
MAF	16q23.1	v-maf musculoaponeurotic fibrosarcoma oncogene	25884094	94753664	90735398–104966351	14 230 953
GALNS	16q24.3	*N*-acetylgalactosamine-6-sulphatase precursor	26701819	107082132	Nearest 106410182	—
CYBA	16q24.3	Flavocytochrome b-558 alpha polypeptide	1535	106941263	Nearest 106410182	—
GFAP	17q21.31	Glial fibrillary acidic protein	20812670	47569101	39868543–48411175	8 542 632
ERN1	17q23.3	Endoplasmic reticulum to nucleus signalling	646932081	70098037	Nearest 70040648	—
CTAGE-1	18q11.1	Cutaneous T-cell lymphoma-associated antigen	100064693	20577869	13392618–54104840	40 712 222
CDH2	18q12.2	Cadherin	1630	28317343	13392618–54104840	See above
DCC	18q21.2	Deleted in colorectal carcinoma	839831630	57850369	54449609–71373485	16 923 876
NTE	19p13.3	Neuropathy target esterase	872510908	10104831	Nearest 11662191	—
NOTCH3	19p13.12	Notch homolog 3	109084854	19462538	16128820–27052527	10 923 707
SSTK	19p13.11	Serine/threonine protein kinase	85483983	4674766	16128820–27052527	See above
RMP	19q12	Transcription modulating factor	81938725	37837413	28221927–44764587	16 542 660
NEUD4	19q13.12	Zinc-finger; neural specific	48588193	47104806	46070350–52764717	6 694 367
NOVA2	19q13.31	Neuro-oncological ventral antigen 2	298444858	57148580	53918418–59278949	5 360 531
TFPT	19q13.34	TCF3 (E2A) fusion partner	533529844	66371798	62726350–67370837	4 644 487
PLCG1	20q12	Phospholipase C, gamma 1	51215335	41433242	28921184–48156350	19 235 166
PCP4	21q22.2	Purkinje cell protein	5121	38093562	Nearest 12292280	—

These data are based on the Human Working draft of August 2001. The sum of all basepairs that are localised between two nearest nonhomologous intrachromosomal duplicons is 734 106 358 bp. This resembles 24.5% of the total human genome. The regions at the chromosome ends are not included in this calculation, as these are not localised between two nonhomologous intrachromosomal duplicons.