Chapter 2
Copy Number Variations

Jianguo Lu and Zhanjiang (John) Liu

Copy number variation (CNV) is a segment of DNA with copy number differences by comparison of two or more genomes. The segment may vary in size, ranging from one kilobase (kb) to several megabases (Cook and Scherer, 2008). Although copy number differences involving segments smaller than 1 kb can also be technically viewed as CNVs, research methods and applications involved in the study of smaller tandem segments are quite different. Therefore, they are not included in the CNV discussions here (see Table 2.1).

CNVs can be caused by changes in genomic architecture including deletions, insertions, and duplications. Low copy repeats are regionally specific repeat sequences, which are susceptible to genomic rearrangements that result in CNVs. The size, sequence similarity, orientation, and distance between the copies of repeated sequences are important factors for causing CNVs (Lee and Lupski, 2006). In spite of being known for a long time, serious research on CNVs and their impact on genomes and genome expression has been a recent event. Structure variations in the human genome have been intensely studied recently. In some cases, the human structural variations in copy number and translocations and rearrangements were found to be associated with disease (Iafrate et al., 2004; Sebat et al., 2004; Tuzun et al., 2005; Redon et al., 2006; de Smith et al., 2007). As the information on CNV accumulates, it is clear that CNVs are important in terms of genome expression and function.

In aquaculture species, research on CNVs is essentially lacking, but it is important to understand the impact of CNVs and its significance in aquaculture. In particular, this is because many aquaculture species are teleost fish. In teleost fish, another major mechanism could also account for many instances of CNV: Teleost fish went through an additional round of whole genome duplication followed by gene loss during evolution. This would mean that teleost fish contain genomes with various levels of CNVs involving coding genes, ranging from almost a complete tetraploid fish to almost a diploid fish with duplicated genes in small proportions of the genomes.

Although CNVs can influence genome expression and function, it is almost certain that CNV polymorphism can influence performance traits, and therefore are highly relevant when discussing whole genome selection. However, as CNV research in aquaculture species is in its infancy, we will not be able to provide information concerning applications of CNV in whole genome-based selection. Rather, in this

Next Generation Sequencing and Whole Genome Selection in Aquaculture Edited by Zhanjiang (John) Liu
 ISBN: 978-0-813-80637-2

Table 2.1 Methods summary for the detection of structural variations in the human genome.

Types	Definitions	References
Single-nucleotide polymorphism (SNP)	Base substitution involving only a single nucleotide	Gibbs et al. (2003)
Structural variant	A genomic alternation (e.g., a CNV, an insertion) that involves segments of DNA > 1 kb	Feuk et al. (2006)
Duplication	A duplicated genomic segment >1 kb in length with >90% similarity between copies	Feuk et al. (2006)
Indels	Variation from insertion or deletion event involving <1 kb of DNA	Feuk et al. (2006)
Intermediate-sized structural variant (ISV)	A structural variant that is ~8–40 kb in size; this can refer to a CNV or a balanced structural rearrangement	Tuzun et al. (2005)
Low copy repeat (LCR)	Similar to segmental duplication	Lupski (1998)
Multisite variant (MSV)	Complex polymorphic variation that is neither a PSV nor a SNP	Fredman et al. (2004)
Paralogous sequence variant (PSV)	Sequence difference between duplicated copies (paralogs)	Eichler (2001)
Segmental duplication	Duplicated region ranging from 1 kb upward with a sequence identity of >90%	Eichler (2001); Sharp et al. (2005)
Interchromosomal duplication	Duplications distributed among nonhomologous chromosomes	Eichler (2001)
Intrachromosomal duplication	Duplications restricted to a single chromosome	Eichler (2001)
Copy number variant (CNV)	A segment of DNA that is 1 kb or larger and is present at a variable copy number in comparison with a reference genome	Feuk et al. (2006)
Copy number polymorphism (CNP)	A CNV that occurs in more than 1% of the population; originally, this definition was used to refer to all CNVs	Sebat et al. (2004)
Inversion	A segment of DNA that is reversed in orientation with respect to the rest of the chromosome; pericentirc inversions include the centromere, whereas paracentric inversion do not	Feuk et al. (2006)
Translocation and rearrangement	A change in position of a chromosomal segment within a genome that involves no change to the total DNA content Translocations can be intra- or inter-chromosomal	Feuk et al. (2006)

chapter, we will provide an introduction of CNV research, summarize methods for CNV discovery, review the different approaches for CNV detection, and discuss the potential application of CNV for aquaculture genome research.

Characteristics of CNVs

As the term CNV itself suggests, it refers to any changes in chromosome structure resulting in the change of copy number—including insertion or deletion of segments in some genomes in the population of a species, but not in all genomes of the population—translocations or rearrangements joining two formerly separated DNA sequences, leading to the net difference in copy numbers among genomes. However, for scientific communications, the definition of CNVs has been quite dynamic. The original definition is that CNVs are intra- or interchromosomal duplications or deletions of segments larger than 1 kb, but not including high copy number repetitive sequences such as long interspersed nucleotide elements (LINEs) or pericentromeric tandemly repeated DNA sequences (Feuk et al., 2006; Freeman et al., 2006). However, CNVs smaller than 1 kb and complex structures within these CNVs among humans have been reported using high-resolution genome maps (Korbel et al., 2007; Kidd et al., 2008). Thousands of insertion and deletion polymorphisms, less than 1 kb in length, have been detected and also referred to as CNVs (Mills et al., 2006). Hence, the broad-sense definition of CNVs is often expanded to include gains and losses of DNA segments of a few hundred bases and larger (Gokcumen and Lee, 2009). It was suggested in that CNVs should not cover insertion/deletion of transposable elements in order to reduce the complexity of CNV analysis ("The Effects of Genomic Structural Variation on Gene Expression and Human Disease Workshop," The Wellcome Trust Sanger Institute, Hinxton, UK, November 27–28, 2005). Therefore, CNVs include copy number polymorphisms (CNPs; Sebat et al., 2004), large-scale copy number variants (LCVs; Iafrate et al., 2004), and intermediate-sized variants (ISVs; Tuzun et al., 2005), but does not encompass retroposon insertions (Table 2.1).

Impact of CNVs on Gene Expression and Phenotypes

CNVs are a vital source in evolution, and have been found to be involved in many human diseases such as developmental disorders, mental diseases, and cancer. In human population, CNV was shown to represent a major type of polymorphism; approximately 12% of the human genome is subjected to CNV (Redon et al., 2006). Although the extent of CNV is unknown at present from many other species, it is reasonable to assume that CNVs are a huge source of genome variation in most, if not all, species. In teleost fish species, due to the additional round of genome duplication followed by gene loss, CNVs could be one of the largest genome variations, and their impact on phenotypes could be tremendous.

CNVs can come from meiotic division processes and somatic division processes. While the meiotic origin of CNVs is well documented, good examples of CNVs derived from somatic processes exist. For example, monozygotic twins (identical

twins) display different DNA CNV profiles (Bruder et al., 2008); CNVs even vary in differentiated human tissues and organs from the same individual (Piotrowski et al., 2008), both demonstrating the mitotic origin of CNVs.

CNV can have great phenotypic impact by adjusting gene dosage, disturbing coding sequence, or regulating long-range gene expression (Kleinjan and van Heyningen, 2005). Gene expression levels can be positively correlated with copy number increment (Somerville et al., 2005; McCarroll et al., 2006) or negatively correlated with copy number increment (Lee et al., 2006). For example, the deletion of a transcriptional repressor can increase gene expression. There is at least 17.7% heritable variation in gene expression caused by CNVs in human (Stranger et al., 2007). Gain and loss of gene functions can have both beneficiary and detrimental impact to the organism. This is particularly true for dosage-sensitive genes.

Most CNV research has been, to date, conducted in humans. However, due to the importance CNV polymorphism, CNV research has recently been conducted in other species, including agriculturally important species such as cattle (Liu et al., 2008, 2010), and chicken (Völker et al., 2010). As CNV research in agricultural species is still at its early stages, phenotypic impact of CNVs awaits further elucidations.

Methods for CNV Detection and Analysis

Microscopic Level Analysis of Structural Variation

CNVs can be detected at the microscopic level through karyotyping. At the molecular level, CNV can be caused by translocations, inversions, deletions, and duplications. If the involved chromosomal segments are large in size, such chromosomal alterations can be detected by cytogenetic techniques such as karyotype analysis. With the improved chromosome banding techniques, many structural variations and structural abnormalities have been identified, especially in disease samples (Jacobs et al., 1978, 1992; Coco and Penchaszadeh, 1982; Warburton, 1991; Barber et al., 1998; Kim et al., 1999). Moreover, with fluorescence *in situ* hybridization (FISH), structural variations can be discerned even when a small chromosomal segment is involved. Chromosome banding also allows detection of a great variety of heteromorphisms. The most commonly detected heteromorphisms involved increases in length or inversions in human chromosome 9 (Verma et al., 1978). The structure variations of this region may involve unequal exchanges and repetitive sequences at recombination positions near the centromere (Starke et al., 2002). It should be noted that cytogenetic techniques such as chromosome banding or FISH has the ability to detect CNVs, but nonetheless, they are insensitive techniques, and they lack the ability to detect genome-wide CNVs of various sizes.

Array Comparative Genome Hybridization (CGH)

In recent years, a number of experimental approaches and computational strategies were used to detect human genome structural variations with different resolutions

(Table 2.1). The most popular approach for the analysis of CNV is the array-based CGH (array CGH or a-CGH).

Array CGH is also called molecular karyotyping. It is a technique to scan the genome for gains and losses of chromosomal segments to discover CNVs (Solinas-Toldo et al., 1997; Pinkel et al., 1998; Lucito et al., 2003; Iafrate et al., 2004; Sebat et al., 2004; Selzer et al., 2005; Tyson et al., 2005). It is a hybridization-based approach using array as a platform. The use of array allowed for the placement of a large number of features (target sequences, sometimes also referred to as probes; but we will use features to avoid confusion with hybridization probes), which in this case are short sequence oligos based on the reference genome sequence. For instance, if the genome of interest is 1 billion base pairs (bp) in size, various numbers of features can be designed to provide the desired resolution. For instance, if one would like to know the copy number situation across the entire genome with one feature every 100 kb, a total of 10,000 short oligos would be required, with each of them designed based on the reference genome sequence with a spacing of 100 kb among them. Array CGH is the most widely used approach for the analysis of CNVs.

The first step of making an array CGH is to place short oligo features representing very short genomic DNA segments spanning the entire genome on arrays (sometimes also referred to as microarrays because of the high density of features). The number of probes depends on the level of the resolution. For example, for a genome with a size of 1 billion bp, a set of 10,000 evenly spaced features would allow detection of CNVs at a resolution of one probe per 100 kb. The higher the resolution desired, the more target sequence features are needed. In an ideal situation, if short oligos of 100 bp is used with no spacing among them, 10 million features would cover the entire 1 Gb genome. That would provide a complete "scan" of the entire genome for any possible CNVs. However, practically, it is a balance between the resolution and the cost that dictates the number of features. The more the features are, the greater the resolution, but the more the costs are as well. Most often, an interfeature spacing of 50–100 kb is used. The tens of thousands of features can be derived from gene coding regions or from noncoding regions of the genome, depending on the purpose of the experiments.

The second step is to fluorescently label the genomic DNA from a test sample and a normal reference sample using different fluorophores, for example, Cy3 and Cy5. The idea is that when equal genomic DNA is used from the test and the normal DNA sample, hybridization of the Cy3-labeled (say normal sample) and Cy5-labeled (say test sample) probes will generate equal signals, thereby yellow fluorescence, if there is no CNVs. Upon possession of any CNVs between the normal reference and the test samples, the hybridization signals will not be equal, thereby generating a red or green fluorescence signal, depending on the ratio of Cy3 and Cy5 hybridization signals. If the test sample has more copy numbers, the Cy5 label will generate a stronger signal than the Cy3 label, and therefore the corresponding probes will be red (Figure 2.1).

Genomes often contain highly repetitive elements that interfere with hybridization. In the designing of the features, repetitive elements should be avoided. Nonetheless, the highly repetitive elements in the genome probes can still cause problems. Therefore, hybridization by highly repetitive elements should be blocked by competitive hybridization using nonlabeled repetitive sequences such as COT-1

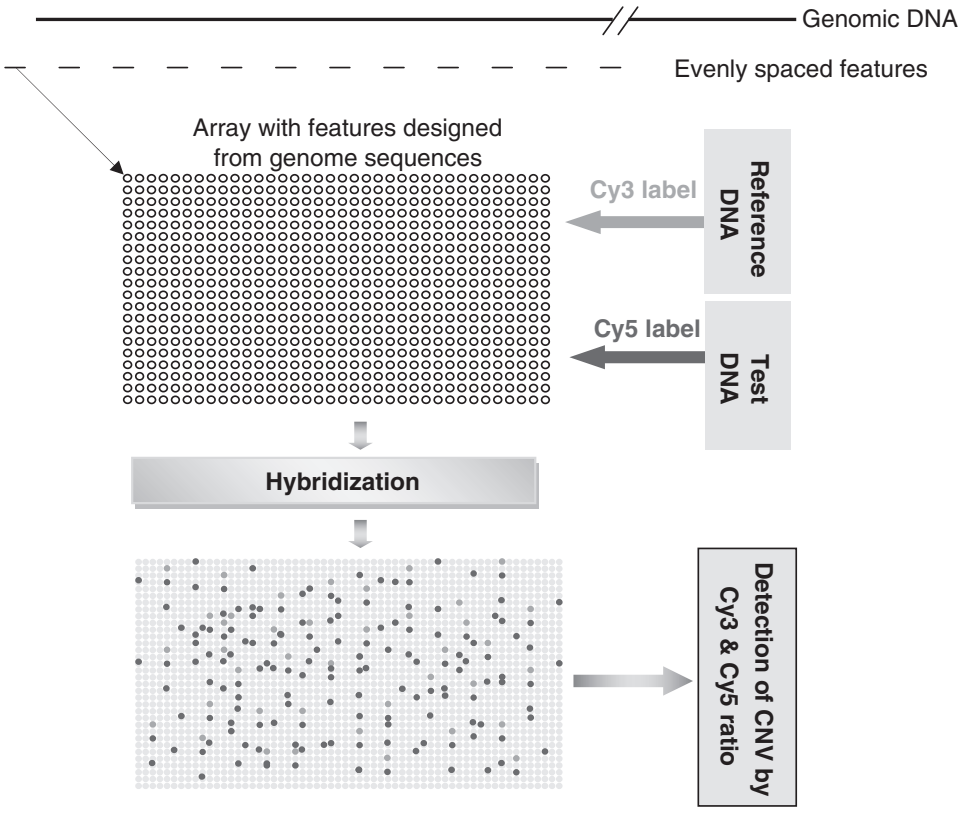

Figure 2.1 Principles of array comparative genome hybridization (array CGH). A large number of evenly spaced features are designed from the reference genome sequence and placed to an array. Equal amount of reference genome (normal genome) and test genome DNA are labeled by differential fluorescence, for example, Cy3 and Cy5, and hybridized to the array. The ratios of Cy3 and Cy5 define CNV. If red fluorescence is observed, the feature on the array has more copy numbers in the test genome than in the normal genome. See color insert.

sequences of human and mouse, which is commercially available. COT-1 DNA is made of highly repetitive sequences based on genome information of the species. The term was derived from reassociation studies using Cot analysis where repetitive DNA reassociates rapidly. COT-1 DNA contains DNA elements with a Cot value of 1.0. In humans, COT-1 DNA is composed of highly repetitive DNA sequences, such as the Alu, LINE-1 and THE repeats. The COT-1 DNA can block the repetitive sequences before the reference and test sample is hybridized to the arrays.

The third step of array CGH is the analysis of hybridization data based on fluorescence ratios. After hybridization, the ratio of the fluorescence intensity of the test probe to that of the reference probe is calculated. The ratio, upon calibration, reveals the copy number differences between the genomes.

The hybridization result can be measured using microarray scanner. Then the feather extraction software can be used to quantify the hybridization images. Finally, the test file outputs are used to do the CNV analysis using CNV detection software.

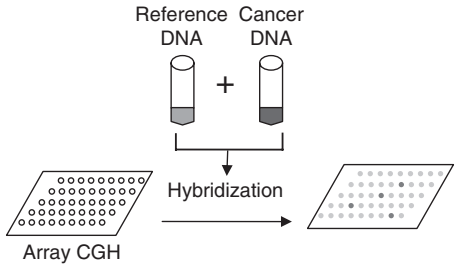

Figure 2.2 An example of using array CGH for the detection of chromosomal segment duplications in cancer. See color insert.

Typical applications of array CGH are for cancer studies because chromosome aberrations usually occur during tumor progression (Albertson et al., 2003) and human genetic disease research (Albertson and Pinkel, 2003; Shaw-Smith et al., 2004). In many cases of cancers, the malignant genome is instable, and segmental duplications can happen in certain genomic regions depending on the cancer type. By using array CGH, it is relatively easy to detect genome regional duplications leading to CNVs (Figure 2.2).

The target sequences on the arrays can be designed based on the needs of the experiments. The targets can be bacterial artificial chromosomes (BACs), cDNAs, polymerase chain reaction (PCR) products, or oligonucleotides (Figure 2.1). The array CGH with BACs have also been widely used recently (Kauraniemi et al., 2001; Ishkanian et al., 2004) because it can provide comprehensive coverage of the genome, low-noise hybridization, reliable mapping data, and accessible clones. However, BACs are usually around 80–200 kb. It is very difficult to detect high-quality single copy number difference smaller than 50 kb, even when hybridization noise is low.

cDNA clones have been used for array CGH to increase the resolution for analysis of single genes or partial genes (Pollack et al., 1999; Kauraniemi et al., 2001; Porkka et al., 2002; Squire et al., 2003). However, there are two shortcomings for this method: (1) the presence of intervening sequences in genomic DNA but not in cDNA due to introns can affect the Cy5 : Cy3 ratio during the hybridization process; and (2) the uneven distribution of genes in the genome (Carter, 2007) would dictate the uneven resolution of the CNV analysis.

Multiplex Amplifiable Probe Hybridization (MAPH)

MAPH is a recently developed procedure for the analysis of CNV in targeted genomic regions based on previously known information (Armour et al., 2000; Patsalis et al., 2005). For instance, certain genes can undergo duplications under malignant tumor conditions. Patient DNA can be subjected to MAPH analysis to detect if the genes are duplicated. In MAPH, target genomic DNA, along with controls in parallel, is immobilized to nylon membranes. Specific genomic segments previously known to be involved in duplications are used as probes. Upon hybridization, and washing away of all unbound probes, the hybridized probes are released and then quantified by

PCR by comparison with the control DNA samples. This method is highly useful for cancer studies, but its application in aquaculture is limited because information of targeted genome duplication is unknown in aquaculture species.

Multiplex Ligation-dependent Probe Amplification (MLPA)

MLPA is another recently developed method for the analysis of CNVs (Schouten et al., 2002). In the MLPA technique, two "half probes" are designed to be adjacent to each other, with each harboring a universal primer sequence linked to its end. Upon hybridization to the target DNA template, the two half probes are brought to proximity to allow ligation. Once ligation happens, a joint molecule is generated that would allow PCR amplification using the known universal primer sequences. The key measurement is the number of the half probes hybridizing to the target. The amounts of ligated probe produced are proportional to the target copy number. Through quantitative PCR of the ligated products, the copy number of the targets is quantified. This approach, in spite of its high specificity, also depends on the prior knowledge of duplicated regions for the design of the half probes. Therefore, its application in aquaculture species is limited. Similar approaches such as quantitative multiplex PCR of short fluorescent fragments (QMPSF) also suffer from the same limitations: requiring prior knowledge for the design of fluorescent primers (Charbonnier et al., 2000), and therefore are not highly useful for aquaculture.

Computational Approach for CNV Detection

Although the CNV detection methods are powerful based on array data, their applications are limited by the array density (array CGH) as well as by costs. CNVs can also be detected based on the computational approach if genome sequences are available. There would be no limitation for the resolution, and the CNVs can be detected at the nucleotide level. The limitation is the unavailability of genomic sequences at present. With the exception of humans, multiple genome sequences are not available from multiple individuals of any species at present. However, with the capacity of next generation sequencing, genome sequences from multiple individuals of the same species will soon become available for many species including agriculturally important species and perhaps even some major aquaculture species.

Paired-end Mapping (PEM) Based on Next Generation Sequencing

A large-scale CNV detection strategy, PEM, was developed recently (Korbel et al., 2007) based on next generation sequencing (Figure 2.3). Basically, with PEM, the genome DNA sequence was first sheared into ~3-kb fragments followed by massive sequencing using next generation sequencing. The 3-kb fragments are ligated to biotinylated adaptors, circularized, and then linearized by shearing. The biotinylated adaptors mark the ends of the genomic fragments and allow the researchers to trace the orientation of the sequences. The sequences generated with next generation

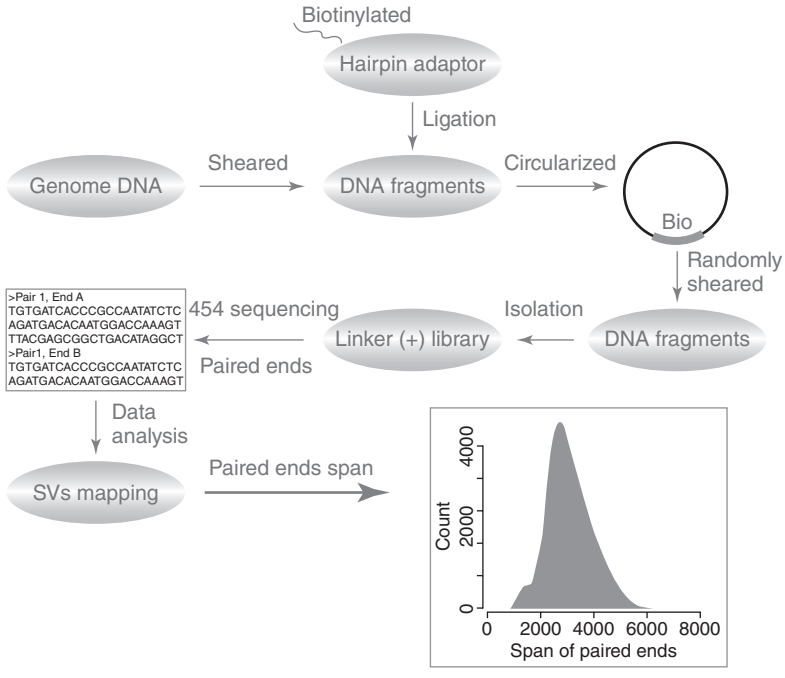

Figure 2.3 Principles of paired-end mapping-based CNV detection. Genomic DNA is sheared into approximately 3-kb fragments. The genomic fragments are then ligated to biotinylated adaptors to mark the orientation. The segments are circularized, followed by linearization at random sites. Next generation sequencing is used to massively sequence the segments. Bioinformatic mapping by *in silico* positioning of the sequences to the reference genome would detect any size difference or orientation difference, which suggest genome structural variations including CNVs. See color insert.

sequencing are then mapped to the reference genome sequence by *in silico* mapping. A deviation in size (maximal 3.0 kb as dictated by the fragment) or orientation (as marked by the orientation of the adaptors) of the sequences generated from paired reads from the reference genome sequence provides evidence of structural variation. If a deletion or insertion is involved, a size difference is expected; if inversion is involved, the sequence orientation is expected to be different (Figure 2.3). Any internal segmental duplication within the sequenced segments would increase the size of the sequenced segments.

Cross-species Computational Analysis

The computational approach is useful for the analysis of gene CNVs among species with reference genome sequence available. For instance, the reference genome sequences are available from zebrafish, medaka, stickleback, fugu, and tetraodon. Analysis of gene CNVs among these would provide perspectives as to how duplicated genomes become diploidized (Lu et al., manuscript in preparation), providing insight into genome evolution.

The advantage of the computational approach for the analysis of CNVs is its great economic benefits without investing large amount of resources if the genome sequences are already available. It also has the advantage of detecting all kinds of structural variations, including translocations, inversions, large-scale CNVs (>50 kb), insertions or deletions (1–50 kb), and small sequence variants (<1 kb) (Feuk et al., 2006). Obviously, computational analysis requires the availability of genome sequences, which is the major disadvantage of computational approaches.

Genome-wide Association Studies Using CNV

The use of genome-wide association studies has successfully linked genetic variants with susceptibility to a wide range of common polygenic diseases. Such genome-wide association studies, however, have almost exclusively focused on single-nucleotide polymorphisms (SNPs). Recent studies, however, have suggested that CNVs may contribute significantly to genetic predisposition to several common diseases (Gonzalez et al., 2005; Aitman et al., 2006). Initially, CNV effect is most often approached through the tagging of CNVs using SNPs, but this approach is not without problems. This has led to the development of association studies directly targeting CNVs. Direct CNV association analysis requires the availability of the maps of common CNV polymorphisms of the genome of interest, which is not available for most aquaculture species at the moment. In the future, it is almost certain that CNV polymorphisms will be found to be important in aquaculture species. The readers are always reminded that the teleost fish have undergone an additional round of genome duplication followed by various levels of diploidization. Therefore, CNVs in teleost fish species will prove to be more important than in other vertebrate species such as humans.

References

Aitman TJ, Dong R, Vyse TJ, et al. 2006. Copy number polymorphism in FCGR3 predisposes to glomerulonephritis in rats and humans. *Nature*, 439:851–855.

Albertson DG and Pinkel D. 2003. Genomic microarrays in human genetic disease and cancer. *Human Molecular Genetics*, 12:R145–R152.

Albertson DG, Collins C, McCormick F, et al. 2003. Chromosome aberrations in solid tumors. *Nature Genetics*, 34:369–376.

Armour JAL, Sismani C, Patsalis PC, et al. 2000. Measurement of locus copy number by hybridisation with amplifiable probes. *Nucleic Acids Research*, 28:605–609.

Barber JCK, Joyce CA, Collinson MN, et al. 1998. Duplication of 8p23.1: A cytogenetic anomaly with no established clinical significance. *Journal of Medical Genetics*, 35:491–496.

Bruder CEG, Piotrowski A, Gijsbers AACJ, et al. 2008. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *American Journal of Human Genetics*, 82:763–771.

Carter NP. 2007. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nature Genetics*, 39:S16–S21.

Charbonnier F, Raux G, Wang Q, et al. 2000. Detection of exon deletions and duplications of the mismatch repair genes in hereditary nonpolyposis colorectal cancer families using

multiplex polymerase chain reaction of short fluorescent fragments. *Cancer Research*, 60:2760–2763.

Coco R and Penchaszadeh VB. 1982. Cytogenetic findings in 200 children with mental retardation and multiple congenital anomalies of unknown cause. *American Journal of Medical Genetics*, 12:155–173.

Cook EH and Scherer SW. 2008. Copy-number variations associated with neuropsychiatric conditions. *Nature*, 455:919–923.

Eichler EE. 2001. Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends in Genetics*, 17:661–669.

The Effects of Genomic Structural Variation on Gene Expression and Human Disease Workshop, The Wellcome Trust Sanger Institute, Hinxton, UK, November 27–28, 2005.

Feuk L, Carson AR, and Scherer SW. 2006. Structural variation in the human genome. *Nature Reviews Genetics*, 7:85–97.

Fredman D, White SJ, Potter S, et al. 2004. Complex SNP-related sequence variation in segmental genome duplications. *Nature Genetics*, 36:861–866.

Freeman JL, Perry GH, Feuk L, et al. 2006. Copy number variation: New insights in genome diversity. *Genome Research*, 16:949–961.

Gibbs RA, Belmont JW, Hardenbol P, et al. 2003. The International HapMap Project. *Nature*, 426:789–796.

Gokcumen O and Lee C. 2009. Copy number variants (CNVs) in primate species using array-based comparative genomic hybridization. *Methods*, 49:18–25.

Gonzalez E, Kulkarni H, Bolivar H, et al. 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science*, 307:1434–1440.

Iafrate AJ, Feuk L, Rivera MN, et al. 2004. Detection of large-scale variation in the human genome. *Nature Genetics*, 36:949–951.

Ishkanian AS, Malloff CA, Watson SK, et al. 2004. A tiling resolution DNA microarray with complete coverage of the human genome. *Nature Genetics*, 36:299–303.

Jacobs PA, Matsuura JS, Mayer M, et al. 1978. A cytogenetic survey of an institution for the mentally retarded: I. Chromosome abnormalities. *Clinical Genetics*, 13:37–60.

Jacobs PA, Browne C, Gregson N, et al. 1992. Estimates of the frequency of chromosome-abnormalities detectable in unselected newborns using moderate levels of banding. *Journal of Medical Genetics*, 29:103–108.

Jones MR, Maydan JS, Flibotte S, et al. 2007. Oligonucleotide array comparative genomic hybridization (oaCGH) based characterization of genetic deficiencies as an aid to gene mapping in *Caenorhabditis elegans*. *BMC Genomics*, 8:402.

Kauraniemi P, Barlund M, Monni O, et al. 2001. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. *Cancer Research*, 61:8235–8240.

Kidd JM, Cooper GM, Donahue WF, et al. 2008. Mapping and sequencing of structural variation from eight human genomes. *Nature*, 453:56–64.

Kim SS, Jung SC, Kim HJ, et al. 1999. Chromosome abnormalities in a referred population for suspected chromosomal aberrations: A report of 4117 cases. *Journal of Korean Medical Science*, 14:373–376.

Kleinjan DA and van Heyningen V. 2005. Long-range control of gene expression: Emerging mechanisms and disruption in disease. *American Journal of Human Genetics*, 76:8–32.

Korbel JO, Urban AE, Grubert F, et al. 2007. Systematic prediction and validation of breakpoints associated with copy-number variants in the human genome. *Proceedings of the National Academy of Sciences of the United States of America*, 104:10110–10115.

Lee JA and Lupski JR. 2006. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron*, 52:103–121.

Lee JA, Madrid RE, Sperle K, et al. 2006. Spastic paraplegia type 2 associated with axonal neuropathy and apparent PLP1 position effect. *Annals of Neurology*, 59:398–403.

Liu GE, Van Tassel CP, Sonstegard TS, et al. 2008. Detection of germline and somatic copy number variations in cattle. *Developmental Biology*, 132:231–237.

Liu GE, Hou Y, Zhu B, et al. 2010. Analysis of copy number variations among diverse cattle breeds. *Genome Research*, 20:693–703.

Lucito R, Healy J, Alexander J, et al. 2003. Representational oligonucleotide microarray analysis: A high-resolution method to detect genome copy number variation. *Genome Research*, 13:2291–2305.

Lupski JR. 1998. Genomic disorders: Structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends in Genetics*, 14:417–422.

McCarroll SA, Hadnott TN, Perry GH, et al. 2006. Common deletion polymorphisms in the human genome. *Nature Genetics*, 38:86–92.

Mills RE, Luttig CT, Larkins CE, et al. 2006. An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Research*, 16:1182–1190.

Patsalis PC, Kousoulidou L, Sismani C, et al. 2005. MAPH: From gels to microarrays. *European Journal of Medical Genetics*, 48:241–249.

Pinkel D, Seagraves R, Sudar D, et al. 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature Genetics*, 20:207–211.

Piotrowski A, Bruder CEG, Andersson R, et al. 2008. Somatic mosaicism for copy number variation in differentiated human tissues. *Human Mutation*, 29:1118–1124.

Pollack JR, Perou CM, Alizadeh AA, et al. 1999. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genetics*, 23:41–46.

Ponchel F, Toomes C, Bransfield K, et al. 2003. Real-time PCR based on SYBR-Green I fluorescence: an alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions. *BMC Biotechnology*, 3:18.

Porkka K, Saramaki O, Tanner M, et al. 2002. Amplification and overexpression of elongin C gene discovered in prostate cancer by cDNA microarrays. *Laboratory Investigation*, 82:629–637.

Redon R, Ishikawa S, Fitch KR, et al. 2006. Global variation in copy number in the human genome. *Nature*, 444:444–454.

Schouten JP, McElgunn CJ, Waaijer R, et al. 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Research*, 30:e57.

Sebat J, Lakshmi B, Troge J, et al. 2004. Large-scale copy number polymorphism in the human genome. *Science*, 305:525–528.

Selzer RR, Richmond TA, Pofahl NJ, et al. 2005. Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosomes & Cancer*, 44:305–319.

Sharp AJ, Locke DP, McGrath SD, et al. 2005. Segmental duplications and copy-number variation in the human genome. *American Journal of Human Genetics*, 77:78–88.

Shaw-Smith C, Redon R, Rickman L, et al. 2004. Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. *Journal of Medical Genetics*, 41:241–248.

de Smith AJ, Tsalenko A, Sampas N, et al. 2007. Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males: Implications for association studies of complex diseases. *Human Molecular Genetics*, 16:2783–2794.

Solinas-Toldo S, Lampel S, Stilgenbauer S, et al. 1997. Matrix-based comparative genomic hybridization: Biochips to screen for genomic imbalances. *Genes Chromosomes & Cancer*, 20:399–407.

Somerville MJ, Mervis CB, Young EJ, et al. 2005. Severe expressive-language delay related to duplication of the Williams-Beuren locus. *New England Journal of Medicine*, 353:1694–1701.

Squire JA, Pei JM, Marrano P, et al. 2003. High-resolution mapping of amplifications and deletions in pediatric osteosarcoma by use of CGH analysis of cDNA microarrays. *Genes Chromosomes & Cancer*, 38:215–225.

Starke H, Seidel J, Henn W, et al. 2002. Homologous sequences at human chromosome 9 bands p12 and q13-21.1 are involved in different patterns of pericentric rearrangements. *European Journal of Human Genetics*, 10:790–800.

Stranger BE, Forrest MS, Dunning M, et al. 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*, 315:848–853.

Tuzun E, Sharp AJ, Bailey JA, et al. 2005. Fine-scale structural variation of the human genome. *Nature Genetics*, 37:727–732.

Tyson C, Harvard C, Locker R, et al. 2005. Submicroscopic deletions and duplications in individuals with intellectual disability detected by array-CGH. *American Journal of Medical Genetics Part A*, 139A:173–185.

Verma RS, Dosik H, and Lubs HA. 1978. Size and pericentric inversion heteromorphisms of secondary constriction regions (H) of chromosomes 1, 9, and 16 as detected by Cbg technique in Caucasians—Classification, frequencies, and incidence. *American Journal of Medical Genetics*, 2:331–339.

Völker M, Backström N, Skinner BM, et al. 2010. Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. *Genome Research*, 20:503–511.

Warburton D. 1991. De-novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal-diagnosis—Frequency, clinical-significance and distribution of breakpoints. *American Journal of Human Genetics*, 49:18–18.