Comparison of multiplex ligation-dependent probe amplification and real-time PCR accuracy for gene copy number quantification using the β-defensin locus

Andrea Perne1*, XiangHong Zhang1*, Lutz E. Lehmann2, Marco Groth3, Frank Stuber4, and Malte Book2

1Department of Anaesthesiology and Intensive Care Medicine, University Hospital Bonn, Bonn, Germany, 2University Department of Anaesthesiology and Pain Therapy, Inselspital, Bern University Department, Bern, Switzerland, and 3Gene Analysis, Leibniz Institute for Age Research–Fritz Lipmann Institute, Jena, Germany

Keywords: β-defensin; copy number variation; real-time PCR; multiplex ligation-dependent probe amplification

Supplementary material for this article is available at www.BioTechniques.com/article/113300.
* A.P. and X.-H.Z. contributed equally to this work.

The reliable quantification of gene copy number variations is a precondition for future investigations regarding their functional relevance. To date, there is no generally accepted gold standard method for copy number quantification, and methods in current use have given inconsistent results in selected cohorts. In this study, we compare two methods for copy number quantification. β-defensin gene copy numbers were determined in parallel in 80 genomic DNA samples by real-time PCR and multiplex ligation-dependent probe amplification (MLPA). The pyrosequencing-based paralog ratio test (PPRT) was used as a standard of comparison in 79 out of 80 samples. Real-time PCR and MLPA results confirmed concordant DEFB4, DEFB103A, and DEFB104A copy numbers within samples. These two methods showed identical results in 32 out of 80 samples; 29 of these 32 samples comprised four or fewer copies. The coefficient of variation of MLPA is lower compared with PCR. In addition, the consistency between MLPA and PPRT is higher than either PCR/MLPA or PCR/PPRT consistency. In summary, these results suggest that MLPA is superior to real-time PCR in β-defensin copy number quantification.

Introduction

Gene copy number variations (CNVs) contribute significantly to interindividual genetic variability (1). One important site of genomic rearrangement is the defensin gene locus located at chromosome 8p23.1 (2). Defensins are antimicrobial and immunomodulating polypeptides that play an important role in the innate immune response, quantitative variations in the gene copy number might contribute to susceptibility to infectious and inflammatory diseases (10).

In principle, CNVs are detectable using screening techniques such as DNA arrays or sequencing. Redon et al. compared the single-nucleotide polymorphism genotyping array method and the comparative genomic hybridization array method for CNV detection (1), finding that both high-throughput methods detected deletions and duplications qualitatively, but failed to generate definitive gene copy numbers. Recently, parallel sequencing emerged as a method for quantitative identification of genomic structural variance, including deletions, duplications, and rearrangements (11,12). However, this method is limited by its use of random amplification primers and lack of definitive copy number quantification. Although genome-wide CNV mapping has been simplified by these different direct screening methods, none of these approaches currently offers reliable single gene quantification.

Consequently, defensin CNVs are analyzed with methods that offer single gene quantification, such as MLPA, PPRT, fluorescence in situ hybridization (FISH), or real-time PCR. To date, a gold standard method for defensin gene quantification has not been established (2,9,13–15). Hollox et al. reported concordant gene copy numbers of DEFB4, DEFB103A, and DEFB104A by multiplex amplifiable probe hybridization and semiquantitative fluorescence in situ hybridization (2). Our own multiplex ligation-dependent probe amplification (MLPA) results confirmed this concordance (14), whereas PCR analysis detected discordance in intra-individual β-defensin copy numbers (13). Other groups have reported inter-method quantification comparisons that showed inconsistent results for absolute copy number (16,17). Therefore, a thorough reevaluation of different quantification methods is required. The aim of this study was to compare MLPA and real-time PCR for β-defensin gene quantification through the analysis of 80 genomic DNA samples.

Materials and methods

Genomic DNA

DNA samples from cell lines and whole blood samples were included. 42 genomic DNA samples, derived from immor-
talized B-lymphoblastoid cells lines, were purchased from the Coriell Institute for Medical Research (Camden, NJ, USA) and the European Collection of Cell Cultures (ECACC; Wiltshire, UK). Samples were named NAxxxx (Coriell) and Cxxxx (ECACC) and have been designated with a superscripted ‘HM’ if they were used in the HapMap Project (18). In addition, 38 DNA samples (code RC) from a white population extracted from whole blood were also analyzed. Written informed consent from healthy volunteers was obtained, as requested by the local ethics committee. This procedure complied with all relevant laws, guidelines, and policies. For simplicity, all specimens will be referred to hereafter as “DNA samples.” The concentration and purity of DNA was determined using a NanoDrop device (NanoDrop Technologies, Wilmington, DE, USA). MLPA and PPRT data from the Coriell and ECACC samples were part of a previous report (14).

**Real-time PCR assay**

The primers and hybridization probes were synthesized according to our previously published approach (13). The albumin gene (ALB) served as a single copy per chromosome reference gene in this study. The method of real-time quantitative PCR using LightCycler Relative Quantification Software 1.0 (Roche, Mannheim, Germany) for analysis requires an appropriate calibrator. Genomic DNA sample NA18608 HM, with one copy of target gene per haploid genome as determined by different methods (14,19) and one copy of reference gene per haploid genome, was used as calibrator. The PCR reaction was performed according to the protocol established by our group (13). In brief, this method of relative real-time quantitative PCR requires a coefficiency file from which the efficiency of PCR is calculated by the software. To create a coefficiency file for the ALB gene and the DEFB4, DEFB103A, or DEFB104A gene, an unknown genomic DNA sample was prepared in a 10-step dilution series. Then, for each dilution, the β-defensin and the ALB genes were amplified in a single capillary to acquire the relative quantification standard curves for both of these genes. Using the LightCycler software, a coefficiency file can be prepared from the two relative quantification standard curves. Data analysis of the real-time PCR method has been described in detail by our group in a previous publication (13).

**MLPA and analysis**

MLPA was carried out using the MLPA Kit P139 (MRC Holland, The Netherlands) as described in our previous publication (14). This MLPA assay uses a total of 43 probes. Ten of these probes hybridize to single-copy genes not located on chromosome eight. The remaining 33 probes are mapped exclusively to chromosome eight, which carries the α- and β-defensin gene cluster. Of these 33 probes, 10 cover the β-defensin cluster, 10 hybridize to genes within the α-defensin cluster, and 13 are specific for genes flanking...
the α- or β-defensin cluster. Concordance analysis was performed using a correlation matrix, and absolute β-defensin copy numbers for each individual were determined using the relative locus dose calculation (14).

### Pyrosequencing-based paralog ratio test

Of the 80 samples, 79 were successfully analyzed by PPRT (one sample was excluded due to failed analysis). The PPRT method is based on the paralog ratio test (PRT) technique described by Armour et al. (17), which targets paralogous gene loci of HSPDP3. The detailed procedures have been described by our group (14). In brief, a first PCR was carried out using unlabeled forward and reverse primers as reported previously (17). Subsequently, a second PCR amplification was performed using the product from the first PCR as a template to obtain biotin-labeled PCR product. This second PCR was set up using the previous forward primer (but 5′-biotin-labeled) and a second reverse primer. The resulting PCR product was used for the pyrosequencing reaction, with the second PCR reverse primer used as the sequencing primer. Two paralog sequence variations (PSV) were used to quantify chromosome 8 and chromosome 5 paralogs in only one pyrosequencing reaction. Pyrosequencing results were rejected if the deviation between the two PSVs was >5%. The copy number (cn) of the chromosome 8 paralog was calculated subsequently using the following function: \[ cn = 2 \times AV8 / (100 – AV8) \], where AV8 describes the average percentage of the chromosome 8 PSV alleles. The calculated values were rounded to the nearest integer.

### Statistical methods

The analysis of the intra-method variances was performed by one-way analysis of variance with Bonferroni’s multiple comparison test. A \( p \)-value <0.05 was regarded as statistically significant. Bland-Altman plots were used to visually assess agreement between the three methods (20). The linearly weighted kappa statistic was calculated to estimate agreement between the methods. Confidence intervals were calculated using bias-corrected bootstrap estimations.

### Results and discussion

### Identification of five gene copy number clusters by real-time PCR

β-defensin gene (DEFB4, DEFB103A, and DEFB104A) quantification was performed in 80 genomic DNA samples by real-time PCR; Figure 1 shows the raw, unrounded copy numbers obtained by this method. The samples were grouped by the mean of the intra-sample normalized ratios for DEFB4, DEFB103A, and DEFB104A (ratios are provided in Supplementary Table S1). This clustering enabled the final copy number identification and resulted in five clusters with two to six gene copies, respectively. The coefficient of variation (CV) in the two-copy cluster was 75% and increased to a maximum of 10% in the six-copy cluster. In 27 out of 29 samples with two or three DEFB4 gene copies, the DEFB103A and DEFB104A gene copy numbers were concordant with DEFB4. In contrast, in the samples with four or more DEFB4 copies, DEFB4, DEFB103A, and DEFB104A gene copy number concordance was detected in only 33 out of 51 samples. Overall, 60 out of 80 samples showed concordant results. In the 20 discordant samples, 19 showed an absolute difference of one copy number, whereas one sample (NA185029M) showed a two-copy number difference (Supplementary Table S1).

### Comparison of copy number quantification methods

The inter-method comparison was performed by using the unrounded β-defensin gene copy numbers detected by PCR and the locus dose-based copy number determination by MLPA. In the absence of a gold standard method, unrounded copy numbers determined by PPRT were used to evaluate the real-time PCR and MLPA results.

Figure 2 shows the β-defensin gene copy number determinations (the numerical values are available in Supplementary Table S2). Real-time PCR and MLPA quantification gave consistent results for copy number for 32 out of 80 samples. Also, by real-time PCR quantification, 30 of these 32 consistent samples showed fully concordant copy numbers for DEFB4, DEFB103A, and DEFB104A. In 29 of the 32 samples with consistent results between the two methods, there were four or fewer gene copies. This emphasizes that consistent MLPA and PCR results tend to occur in the lower—copy number range and in samples with high intra-sample concordance among the three genes in real-time PCR results.

In the 32 samples with consistent real-time PCR/MLPA results, PPRT confirmed the quantifications in 24. Of note, 51 out of 79 samples showed consistent MLPA and PPRT copy number quantification results, but consistent real-time PCR and PPRT quantification results were detectable in only 35 out of 79 samples. The CV of the raw data was calculated to evaluate the performance of the three methods. The CVs of real-time PCR, MLPA, and PPRT were calculated and grouped by copy numbers (Supplementary Figure S1). One-way analysis of variance with Bonferroni multiple comparison test showed significantly greater CV in the real-time PCR (\( P < 0.05 \)) and PPRT (\( P < 0.05 \)) methods compared with MLPA (Supplementary Table S1).

---

**Next generation sequencing applying the Roche GS FLX TITANIUM Technology**

**Service features:**

- **De novo** sequencing of prokaryotic and small eukaryotic genomes
- Finishing of **de novo** sequencing
- Analysis of metagenomes, transcriptomes / normalised cDNA, methylation patterns, pools of tagged fosmids and BACs
- **Targeted** re-sequencing
- ChiP and ncRNA sequencing

—

**YEAR END OFFER 2009**

Please check our webpage [www.lgc.co.uk/genomics](http://www.lgc.co.uk/genomics) for more information.
enzyme ratio in the range of high copy numbers. A modified amount of template in the high—copy number range might shift the start of the log-linear target gene amplification phase to higher cycle numbers.

In MLPA, hybridization probes hybridize specifically to the target sequence and ligate with the corresponding probe prior to the amplification procedure. The discriminatory power (concordant versus discordant) of the probes used in the present investigation was confirmed by probe to probe correlation coefficients, which were previously published by our group (14). Moreover, incomplete hybridization would not be expected to significantly influence the copy number quantification because a reference gene signal based on the averaged signal intensity of five nearest-neighbor probes is used, and the target gene signal represents the average of 10 gene-specific probes within the 9-defensin locus. MLPA amplification of all target and reference genes was performed in parallel by a single pair of primers. Consequently, compromised reaction efficiency would bias the amplification efficiency of all target sequences equally. This fact might contribute to the lower coefficient of variation of MLPA as well as to its superior discriminatory power in the high—copy number range.

**Evaluation of inter-method consistency**

For visualization and analysis of inter-method agreement, we chose Bland-Altman plots comprising the 79 MLPA, PPRT, and PCR quantified samples (Figure 3). These plots display the intra-sample differences against the intra-sample mean, and are based on the assumption that the mean of the two measurements is the best available estimation (20). The limit of agreement was calculated as two times the standard deviation of the differences. The smallest limit of agreement was found in the MLPA/PPRT comparison (range -1.0–1.7; Figure 3C), whereas the comparison of real-time PCR with MLPA (Figure 3A) and PPRT (Figure 3B) resulted in higher ranges (-2.6–0.9 and -2.5–1.5, respectively). Thus, the Bland-Altman plots indicate the limited consistency of real-time PCR with both of the other methods. However, both plots show higher consistency in the low—copy-number range compared with the high—copy number range. This observation is consistent with the decreased concordance seen within the high—copy number PCR results (Figure 1).

Although Bland et al. suggested that methods yielding values within the limits of agreement can be considered clinically interchangeable (20), this conclusion is not applicable to gene copy number quantification, where a reliable quantification is mandatory. However, in the absence of a gold standard method, the assessment of agreement by these plots is an important contribution to the evaluation of different methods.

Additionally, the estimated agreement between the methods was calculated by linearly weighted kappa coefficients. The kappa coefficients for inter-method agreement are corrected for random agreement and are displayed in Table 1. According to Landis and Koch (21), the range 0.41–0.60 can be interpreted as moderate agreement, 0.61–0.80 as substantial agreement, and 0.81–1.00 as almost perfect agreement. These data confirmed the results seen in the Bland-Altman plots, in that there is limited consistency between real-time PCR and both alternative methods.

In the end, there is increasing evidence for the functional relevance of copy number variations and, moreover, positive associations with diseases have been reported (2,10,22). Reliable quantification is the basic precondition for further research in the rapidly growing field of genetic variability and predisposition in complex diseases. In this study, we found that two quantification methods—real-time PCR and MLPA—are affected similarly by decreased reliability in the discrimination of high copy numbers. This makes intuitive sense because, presuming optimal conditions, the difference between 2 and 4 gene copies causes an increase of 100% signal amplitude, while the change from 8 to 10 copies results in only a 25% increase in amplitude. Despite this limitation, the presented results indicate the advantage of MLPA compared with real-time PCR in gene copy number quantification.

**Acknowledgments**

The authors thank Sven Trelle from the Clinical Trial Unit of the Inselspital, Bern University Hospital for the excellent statistical advice. This work was supported by a grant from the "Deutsche Forschungsgemeinschaft" (Bo 1929/1-2).

**Competing interests**

The authors declare no competing interests.

**References**


Received 6 July 2009; accepted 9 October 2009.

Address correspondence to Malte Book, University Department of Anaesthesiology and Pain Therapy, Inselspital, CH-3010 Bern, Switzerland. email: malte.book@dkf.unibe.ch