BREEDING AND GENETICS

Chicken Single Nucleotide Polymorphism Identification and Selection for Genetic Mapping[1]

R. Jalving,[2] R. van't Slot, and B. A. van Oost

Department of Animals, Science and Society, Faculty of Veterinary Medicine, Utrecht University, PO Box 80.166, 3508 TD Utrecht, The Netherlands

ABSTRACT Single nucleotide polymorphisms (SNP) are the ideal markers for high-density genome wide mapping. A total of 327,000 expressed sequence tag (EST) sequences, obtained from the ChickEST project, were examined for the presence of SNP. A total of 32,268 potential chicken SNP were identified and stored in a customized Microsoft Access database and evaluated in silico for their usability for a high-density genetic map. Based on a minimum of 3 for the minor allele occurrence and a minimum of 30% for the minor allele frequency, 5,332 reliable SNP were selected, of which both SNP alleles were present in the database at a high frequency. To test the usefulness of the in silico SNP identification, 24 SNP affecting a *Bgl*II site were used for a genotyping study. A functional PCR assay could be designed for 21 of the 24 SNP. It was possible to validate 90% of this marker subset (21 SNP) by *Bgl*II restriction analysis. The high percentage of validated markers demonstrates the reliability of the 5,332 chicken SNP markers. Furthermore, the limited number of genomic DNA samples necessary to validate 90% of the SNP markers confirmed the prediction of the high frequency at which both alleles of the selected SNP were present in the tested chicken populations.

(*Key words*: chicken, single nucleotide polymorphism, genome mapping, selection, high density)

2004 Poultry Science 83:1925–1931

INTRODUCTION

Single nucleotide polymorphisms (SNP) are the most abundant type of polymorphism in the genome. Therefore, SNP are becoming the preferred marker for high-density mapping. In the chicken genome, SNP have been identified with a frequency of 1 SNP per 225 bp, which is 5 times as many as in humans (Vignal et al., 2002). Furthermore, a wide variety of chicken breeds is available (Lamont, 2003) and a well-designed or carefully selected resource population can be obtained within a reasonable time period. Therefore chickens are suitable organisms for the development of high density mapping techniques (Andersson and Georges, 2004). Genome wide analysis using high-density SNP maps will be beneficial to several areas of poultry science. The SNP markers are already used to identify disease resistance genes in chickens (Emara and Kim, 2003; Malek and Lamont, 2003) and can be used as alternatives for microsatellite markers in other chicken-related research topics.

The entire process of data gathering and SNP characterization, which is necessary before any SNP can be used for high-density mapping, involves 4 stages: DNA sequencing, SNP identification, SNP validation, and SNP genotyping. The DNA sequencing, the first stage in SNP characterization, can now be done with high efficiency, the entire process from colony picking to sequence assembly can be done automatically (Meldrum, 2000). To enable automatic sequence assembly, base-calling software, sequence assembly programs, and assembly viewing tools have been created. Phred (Ewing et al., 1998) is a base-calling program, which analyses the sequence chromatogram files from the DNA sequencer, determines the corresponding nucleic acid sequence, and gives an estimate on the reliability of the determined base, the so-called quality value. For assembly of the resulting files, several sequence assembly programs exist, amongst which are Phrap (Green, http://www.phrap.org), Cap3 (Huang and Madan, 1999), and Gap4 (Staden et al., 1998). The produced assemblies can be viewed, analyzed, and edited with Consed (Gordon et al., 1998) and Gap4.

The sequences and, if possible, the corresponding quality values and peak heights, are used in the second part of the SNP characterization process, SNP identification.

Received for publication March 9, 2004.

Accepted for publication August 27, 2004.

[1]The single nucleotide polymorphisms reported in this paper have been submitted to the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/) and have been assigned accession numbers ss28453177 to ss28458508.

[2]To whom correspondence should be addressed: r.jalving@vet.uu.nl.

Abbreviation Key: BBSRC = Biotechnology and Biological Sciences Research Council; EST = expressed sequence tag; SNP = single nucleotide polymorphism; VBA = Visual Basic for Applications.

One approach to identify SNP, which is used by Polyphred (Nickerson et al., 1997), is the comparison of the peak heights of genomic sequence alignments. Heterozygote genomic nucleotides have a reduced peak height for the called base and an additional peak for the uncalled base. A different approach is used by PolyBayes, which uses the quality values to calculate the probability that a sequence difference is a SNP (Marth et al., 1999). However, in the public nucleic databases, peak heights and quality values are often not available and only sequence comparison can be used. Successful attempts to distinguish the SNP from the sequence errors were achieved with the use of a sequence error filtering scheme (Picoult-Newberg et al., 1999), and the use of quality measures based upon frequency and cosegregation of sequence differences (Batley et al., 2003).

For the validation and genotyping of SNP, similar techniques are used. Many different SNP typing technologies are available (Vignal et al., 2002). Most of the currently used methods are so-called single base extension assays. In such an assay a location-specific oligonucleotide is designed adjacent to the SNP of interest. This oligonucelotide is elongated with a SNP allele-specific base and the resulting product is analyzed. Several high-throughput single base extension methods are available, amongst which are MALDI-TOF (matrix-assisted laser desorption-ionization time-of-flight) mass spectrometry analysis (Tang et al., 1999) and the Illumina BeadArray platform (Oliphant et al., 2002). A different methodology based on allele-specific hybridization makes use of a DNA chip containing allele-specific oligonucleotide probes (Wang et al., 1998). For each SNP, 2 oligonucleotide probe arrays are present on the chip, one for each allele. Only a complete match with the present genotype gives a clear signal. High-throughput technologies, such as those mentioned above, make genome-wide marker analysis feasible.

The chicken genome project is in progress and the National Institutes of Health released the first draft sequence on March 1, 2004. As part of the genome project, a physical bacterial artificial chromosome map of the chicken genome has been published (Ren et al., 2003). A consensus linkage map of the chicken genome, containing 2,000 genetic markers, has also been published (Groenen et al., 2000). Several expressed sequence tag (EST) sequence projects have resulted in more than 400,000 EST sequences in the dbEST database as of November 14, 2003 (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html), and the first collection of identified chicken SNP has been reported (Kim et al., 2003). However, a high-density genetic SNP map is not yet available. The aim of the present study was to devise a new SNP selection method and to identify SNP suitable for high-density mapping in the chicken.

MATERIALS AND METHODS

DNA Sequence Data and EST Assembly Tools

The EST sequences used were obtained for the BBSRC (Biotechnology and Biological Sciences Research Council) ChickEST sequencing project (Boardman et al., 2002). The sequences were downloaded from the ChickEST ftp site (ftp://rocky.bms.umist.ac.uk). The sequences in the downloaded FASTA file were assembled with the TGICL script (Pertea et al., 2003). The script is available at the TIGR Web site (http://www.tigr.org/tdb/tgi/software). The resulting ace-file was converted from a Linux text-file to a Windows text-file with the tofrodos utility (http://www.thefreecountry.com/tofrodos/index.shtml).

SNP Identification, Storage, and Analysis

For the identification and analysis of SNP, Visual Basic for Applications (VBA) scripts were developed in Microsoft Access. The VBA scripts were developed for file navigation, FASTA import and export, SNP identification, Primer3 parsing, MegaBlast parsing, trace-file retrieval, Staden Package pregap initiation, automated local blast execution, SNP data analysis, and Microsoft Excel parsing. Microsoft Excel was used to display the distribution of the SNP in concordance to the SNP selection measures. For BLAST analysis, a locally installed version of the NCBI BLAST tools (Altschul et al., 1990) was used (ftp://ftp.ncbi.nlm.nih.gov/blast/executables). The Windows-based BLAST executables of the 2.2.6 release were used. The interface and analysis of the BLAST results were done by a VBA script.

Trace Walking

The SNP of interest were selected and exported as a single FASTA file from the Access database. The presence of introns in the 501 nucleotides of SNP consensus sequence was determined by Megablast analysis using the Gallus_Gallus_WGS_Trace database of the National Center for Biotechnology Information Trace archive (http://www.ncbi.nlm.nih.gov/Traces). For this purpose the multiple FASTA file was taken as query file, filters were switched off, the Expect value was set at 1×10^{-10}, word size was set at 16, and percentage identity was set at 80. The result file was saved as text-file and the Mega Blast parser script of the SNP database retrieved the relevant trace-files. The trace-files were subsequently used to generate a reliable contig in the Staden Package (Staden et al., 1998). If the resulting contigs did not contain enough sequence downstream or upstream of the SNP the terminal traces were used for a similar Megablast procedure to obtain more flanking sequence. This last step was repeated until the contigs were of sufficient length for primer design.

SNP Validation

Selection of SNP for the presence of a *Bgl*II restriction site at the SNP position was done with a Perl script. For this purpose all SNP were exported from the SNP database with both alleles between brackets and 5 upstream and downstream nucleotides. To design primers for selected SNP, an input file for primer3 (Rozen and

TABLE 1. Primers used in the study

Single nucleotide polymorphism	Forward primer	Reverse primer
ss28453780	AAATTGAGTTCATTGGCATC	TTACCTGGATCTTGAGCACT
ss28454323	CTAGGAACGTTTTGTGACATC	GCAATTCCGCTATACAAGAT
ss28454373	ACACAAGCAACCACTTACCT	CGTGTGCACTTAAAAAGACA
ss28454376	TGTTTTTGAGTTGGGAAGAG	CGCTAGCAACAAAGGTACA
ss28454638	AAGTTCTCTTGAACAGTTGACC	ATCACACAGCCTCCAAAG
ss28454701	ACGTGGCATTCCTTCTTT	AGGACAAAGTGCCAGATTT
ss28454719	AGTGAGATTAGGAGCAATGG	GTTAGCCTCCTTGGAATTTT
ss28454626	GACTGAAGACATCCCTAGAATC	ATTAGGTTCTGGCACACAAA
ss28454797	GCTTCTTAACACCCCAAAAT	AGGACCAAACAGACGTATTTC
ss28454911	AGAGCACCTCCATTCCAC	TATGTGCAATCTCAGGACAA
ss28455292	ATGAGATATTGGCCTTGTCA	TTCCTTATGGGGAGAACAC
ss28455294	CTCTGCTGTCTTTCCTTTCTC	AGGGACTCAATATTCCAATG
ss28455382	TGACAAACATCATCAGCCTT	GGTCTTTGAGGAAGGGTTTA
ss28456446	AGCATAGCAGACATTCTTCC	GGTTCAGGTCTGTCTTCTTG
ss28456506	CCACGTTTCTGCAATTTATG	TGCTAGTGACACAAACGAAG
ss28457063	AAGAGCTACAGAGAACTAATGTGAC	GCTACAGGCTTAATTTCACAG
ss28457094	GACTCTTTGAAATCCTGGTG	GAAAATTGCAGGAGTCAGTC
ss28457561	CTTCAGTGCCAACATCTGTA	GGCTGCATACTTGAACCTTA
ss28457579	AAGTGATAGACTGGCCTTGAT	CAAAGAAGGCAACATATACCTC
ss28457780	GTATTTCACCTTGAACTGTGTG	AGGTCTTACCCCATAGATTGTT
ss28458059	ACCGTGAACAGACTTGATTT	GGAAAAACTCCCATTACTGA

Skaletsky, 2000) was generated containing all the relevant parameters (Vieux et al., 2002) by the primer3 parsing script of the SNP database. A locally installed primer3 version (http://www-genome.wi.mit.edu/genome_software/other/primer3.html) was used to design the necessary primers (Table 1).[3] For validation of the SNP, a standard PCR procedure was used to amplify the SNP specific region. The procedure started with heat denaturation for 4 min at 94°C followed by 30 thermal cycles, composed of 30 s denaturation at 94°C, 30 s annealing at 55°C, and 45 s elongation at 72°C. The procedure was concluded with a 15-min elongation at 72°C. After ethanol precipitation of the PCR products, half of the volume of the samples was digested with *Bgl*II and the resulting products were compared with the undigested PCR fragments on a 2% polyacrylamide gel. Twelve chicken genomic DNA samples were used a template. Six of the 12 samples used were full-sib offspring samples of an extreme broiler × broiler cross from a White Plymouth Rock population. The other 6 samples were unrelated White Leghorn samples.

RESULTS

Chicken EST Sequence Assembly

Within the BBSRC ChickEST project, a collection of 64 cDNA libraries, generated from 21 different embryonic and adult tissues, was used to generate EST sequences. As source for the identification of chicken SNP, 327,000 EST sequence reads were retrieved from the BBSRC ChickEST project, and aligned into assemblies, using the TGI clustering tools. This was a 2-step process in which the sequences were first clustered based upon their sequence similarity, after which assembly was performed within each of these clusters. This procedure resulted in an ace-file, containing 25,399 clusters with in total 39,095 contigs and 239,555 sequences.

SNP Identification

For the identification of SNP in the ace-file a program was needed that would be capable of extracting reliable SNP information even without the presence of trace quality values and which could differentiate between highly frequent and rare SNP. In addition, the stored data should be easy to access and to analyze, and the software should be able to run on a standard desktop computer. As none of the existing programs met these criteria, a set of VBA scripts in Microsoft Access was developed for this purpose.

Using a VBA script, the contigs in the ace-file were analyzed one by one to identify potential SNP. Only contigs containing more than 3 reads and a consensus sequence of at least 21 nucleotides were evaluated by the script. For the identification of SNP, analysis of the 20 characters in the direct vicinity of the SNP was sufficient. Therefore, the VBA script started at the 5′ end of each contig and moved toward the 3′ end with a window of 21 characters (Figure 1). During this process, the script evaluated the consensus sequence and the sequence reads to determine the quality of the separate sequences. At the same time it determined if the central character in the window was a potential SNP.

Additional bases or sequence gaps were characteristics of miscalled or misaligned sequence. Additional bases in one or more reads were identified by the presence of asterisks in the consensus sequence. Sequence gaps in the reads resulted in asterisks in these read sequences. Characters in the consensus sequence other than A, C, G, and T, or with more than 2 adjacent asterisks on one side

[3]Isogen, Maarssen, The Netherlands.

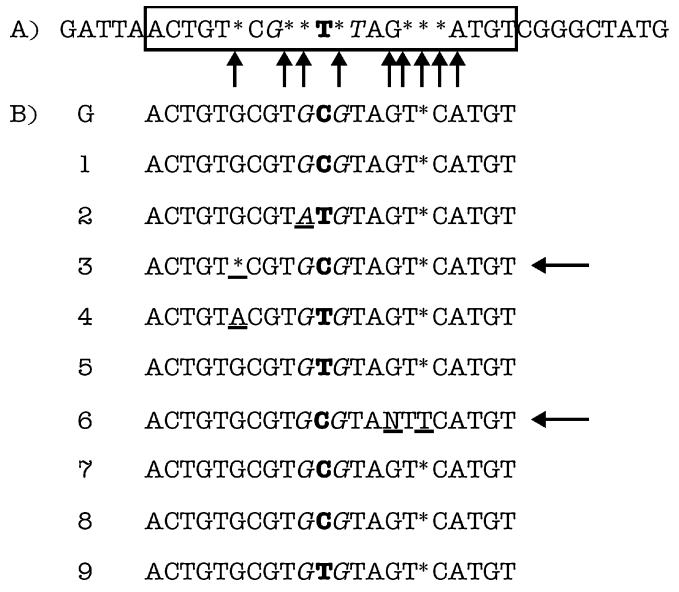

FIGURE 1. A) Consensus character evaluation. A 21-character window (indicated by the box) progresses through the consensus sequence, while the central character (in bold) is evaluated as a possible single nucleotide polymorphism (SNP). The asterisks represent gaps introduced into the consensus as result of extra bases present in a minority of read sequences. Bases in italics are regarded as being adjacent to the bold central base. Arrows indicate unsuitable characters for SNP analysis. B) Read character evaluation. C represents the consensus sequence, to which the reads (1 to 9) are compared. Underlined characters indicate differences between the read and the consensus other than the candidate SNP (bold). The arrows indicate rejected read sequences (due to bad quality sequence). Characters in italics are the adjacent bases of the candidate SNP.

were rejected by the script for SNP identification (Figure 1A). For every remaining character, the corresponding 21 characters of each read of the contig were analyzed in a 3-step procedure.

First, partial sequence reads or sequence reads of bad quality were identified and removed from further analysis. A sequence was qualified as partial if it had less than 10 characters upstream or downstream of the SNP. Sequence reads were considered to be of bad quality if they differed with the consensus sequence in anything other than sequence content, within the 21-character window (Figure 1B).

Every base change could, in theory, be a SNP; however, most base changes were probably sequence errors. Especially in the larger contigs, sequence errors outnumbered the SNP. To avoid selection of misaligned or miscalled bases, a threshold was necessary to correctly identify SNP against the background noise of sequence errors. Therefore as a second step in the procedure, all combinations of the central base with the 2 adjacent bases were counted in the remaining reads. In the hypothetical example shown in Figure 1B, this resulted in 3 GCG (read 1, 7, and 8), 1 ATG (read 2), and 3 GTG (read 4, 5, and 9). If 2 or more combinations were present, a SNP was assigned if the 2 most frequent combinations could be the 2 alleles of a SNP. In addition, both these combinations needed to be present at a higher number than the specified lower threshold. This lower threshold stated that any combination had to be present at least twice. Furthermore, the threshold also stated that if more variations were present and the third most frequent combination was present x times, than the SNP alleles had to be present at least (x + 2) times. In the example of Figure 1B, GCG was equal to the consensus sequence, GTG only differed in the central base, and both could be alleles of a SNP. Because both combinations were present 3 times, and ATG was only present once, a C/T SNP was assigned, with a frequency of 3 Cs and 3 Ts.

Finally, to enable more stringent SNP selection, the quality of the reads used to identify the minor SNP allele (not present in the consensus sequence) was determined as percentage of sequences that did not have additional sequence differences. In Figure 1B, reads 4, 5, and 9 were identified to have the 'minor' SNP allele, read 4 had an additional sequence difference. Therefore, this SNP got a confidence score of 67%.

Using these criteria, 32,268 potential SNP in 9,790 consensus sequences were identified. Both the consensus sequences and the SNP were stored in Microsoft Access. In addition to the nucleotides that defined the SNP, each record of a SNP contained up to 250 nucleotides of sequence of both the upstream and downstream flanking sequence. Moreover, the records contained the frequency of both alleles, the confidence score, the name of the corresponding consensus sequence, and the relative position in that consensus sequence. Of these, 24,255 SNP in 9,376 consensus sequences had a confidence score of 100%. The SNP with a 100% confidence score were analyzed for the presence of duplicates in the Access database with BlastN. In the BlastN analysis, the SNP were used as source for the database and as query, and all SNP that generated a significant hit on a different contig were tagged in the database as having a possible duplicate. After the analysis, 18,168 unique SNP remained.

High-Stringency SNP Selection

To obtain a smaller set of more reliable SNP markers, a second (more stringent) selection was performed on the 18,168 remaining SNP. The aim of this second selection step was to obtain about 5,000 markers, in which both alleles were present at a substantial proportion in the population and would therefore be useful for a genome-wide mapping experiment. The use of a combination of 2 measures, the minor allele frequency and the minor allele occurrence, was examined as more stringent SNP selection criteria. The minor allele frequency was defined as the percentage of sequences representing the minor allele; the minor allele occurrence as the absolute number of sequences that represent the minor allele. The effect of a stepwise increase of these 2 measures was analyzed and applied to the database SNP collection as minimal threshold level for SNP selection (Figure 2). Increasing the minor allele occurrence resulted in a fast decline of the number of SNP, whereas the increase of the minor allele frequency resulted in a more gradual decrease of remaining SNP. Taking selection criteria of 3 for minor

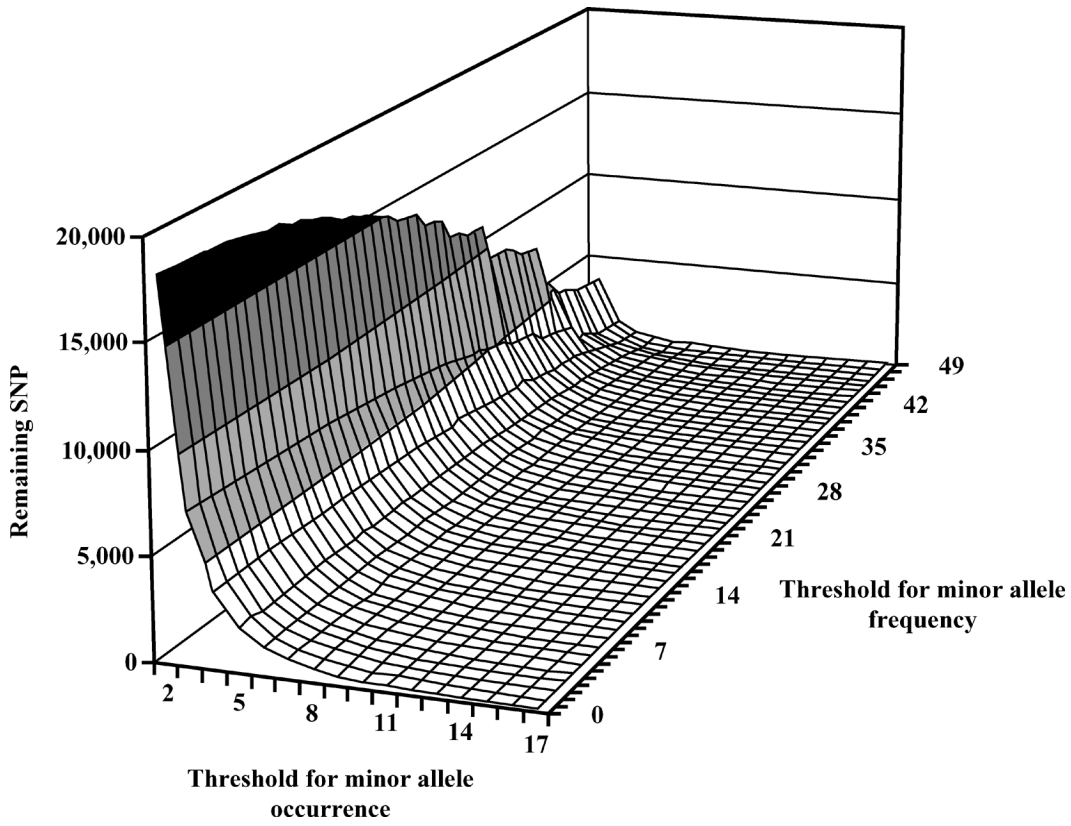

FIGURE 2. Determination of optimal selection criteria. The effect of a stepwise increase of the minor allele occurrence and the minor allele frequency as minimal threshold on the amount of remaining single nucleotide polymorphisms (SNP) is shown. The colored areas indicate the amount of remaining SNP: 1 to 5,000 (white); 5001 to 10,000 (light gray); 10,001 to 15,000 (dark gray); and 15,001 to 20,000 (black).

allele occurrence and a minor allele frequency of 30%, 5,332 SNP were selected for genome-wide mapping purposes.

For validation purposes, the 5,332 SNP were analyzed for the presence of a *Bgl*II restriction site at the SNP position, resulting in a subset of 24 SNP. Megablast analysis was performed with the 501-bp sequences of these 24 SNP, using the traces of the chicken whole genome shotgun sequencing project as database. Intronic sequences, identified by trace walking, were inserted in the 501-bp SNP sequences. Assays were designed for PCR amplification for each of the SNP sequences. Successful assays were designed for 21 SNP. The PCR products were obtained with these 21 primer pairs, using 6 broiler genomic DNA samples and 6 layer genomic DNA samples as templates. The PCR products were digested with *Bgl*II and subsequently compared with the undigested PCR products, to validate the presence of the polymorphism. Replication of the restriction analysis resulted in identical restriction patterns (Figure 3), excluding the possibility that some of the restriction patterns were the result of partial digests. Genotyping of the broiler and layer samples resulted in validation of, respectively, 13 and 16 of the 21 SNP. Nineteen of the 21 SNP were validated in either or both broiler and layer samples. The remaining 2 SNP did not show the polymorphism in these samples. Therefore, using only 12 genomic samples, 90% of the SNP were validated.

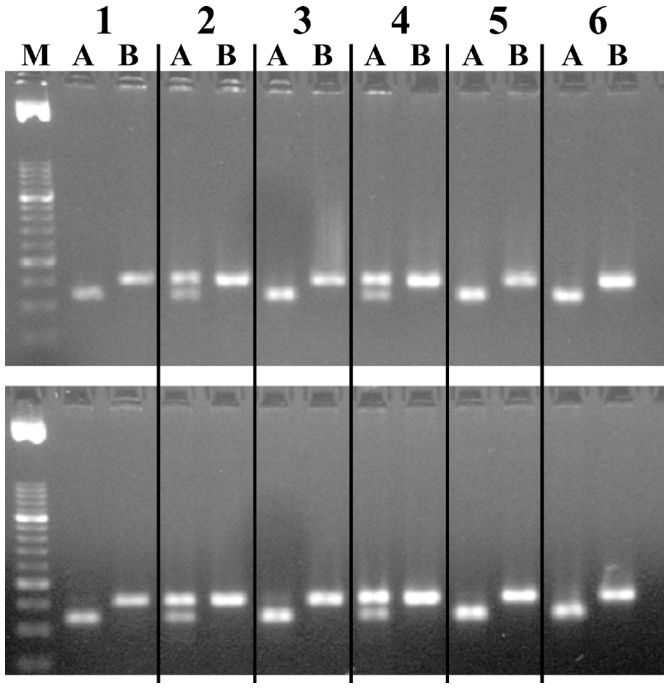

FIGURE 3. Validation of single nucleotide polymorphism (SNP) ss28457561 by *Bgl*II restriction analysis. In 2 independent experiments (upper and lower gel), PCR was performed on 6 genomic samples (1 to 6) to amplify the region in which the candidate SNP was located. The PCR products were digested with *Bgl*II (A) and compared with undigested PCR product (B). Lane M = molecular weight marker (50-bp ladder).

DISCUSSION

The essence of SNP identification is elimination of sequence errors and misaligned bases; therefore, if quality bases are available these should be used. Polyphred was shown to be successful in this approach (Nickerson et al., 1997). For an increasing number of organisms, the quality values of genomic sequencing projects become available and can be obtained from ensemble (http://trace.ensembl.org) or the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/Traces/trace.fcgi). However, trace files are often not available for EST sequences, and EST sequences are a particularly good source for SNP identification (Picoult-Newberg et al., 1999). The available tools for SNP identification in EST sequences aim for maximum reliability of potential SNP, but none of them allow for the flexibility of data analysis and SNP selection that was desired. Therefore, a new SNP identification tool was developed which aimed for reliable SNP identification and which offered flexible SNP selection to obtain a SNP subset, based on selection criteria for the planned experiment.

For the development of any software application it is important to keep in mind the purpose of the tool and choose the programming language best suited to develop the application. In this case it was important to have maximum programming flexibility (scripting language) and good data storage capabilities. After the SNP identification, it should be possible to perform easy and fast data analysis. Visual Basic for Applications in the Microsoft Access database program fulfilled all these requirements, and was used for the development of the new SNP identification tool.

The selection criteria for SNP identification depend on the field of interest. For the identification of genetic variation it is important to exclude all possible sequence errors. For the identification of rare SNP, all possible SNP need to be identified, and for the development of a SNP map, the objective is to obtain the most likely SNP with a high frequency. Therefore, in the first stage of SNP identification, maximum reliability of the SNP was not the aim. The aim was to exclude obvious or very likely sequence errors, and store the data, which was needed to perform interest-specific selection.

Without base quality information, all indications of miscalled or misaligned bases have to be obtained from the aligned read sequences. Bad sequence quality can be recognized by the presence of sequence differences between the different reads in a contig but also by the presence of additional bases or sequence gaps in one or more reads. Sequence reads that contain additional bases or additional gaps are not high quality sequences and should be removed from further analysis. Sequence differences, another characteristic of bad sequence quality, can also be due to genetic variation. Apparently, the SNP density varies a lot within the chicken genome and more than one SNP can be present in the 21-character window. Removing all sequence reads that have additional sequence differences was therefore not opted for, because it would result in the removal of several SNP in the first step of the identification. It is therefore important to determine which sequence differences are miscalled bases and which sequence differences are SNP. The selection method that was used counted how often the alleles of an SNP were present, and if this was substantially more than the amount of possible sequence errors present in the same genomic region, the SNP was entered in the database.

The database offers the possibility for a subset selection to obtain SNP best suited for the field of interest. To make this second selection procedure possible, a confidence score was determined together with the SNP that allowed for a maximum reliability selection. Two other measures that can be used for the second phase selection are the minor allele occurrence and minor allele frequency. To calculate those 2 measures, the presence of both SNP alleles in the contig is added to the SNP information in the database.

The most useful SNP for a mapping experiment are reliable SNP of which both alleles are present at a high frequency, and this was therefore used in the second selection procedure. To increase the reliability of the selected SNP, the confidence score and the minor allele occurrence were used as measures. To ensure a high frequency, the minor SNP frequency was used as measure. The selected subset contains 5,332 SNP.

Twenty-one SNP were tested in 12 chicken genomic DNA samples to determine the reliability of the identified 5,332 putative SNP. Ninety percent of the tested SNP were validated. As expected, more SNP were validated in the White Leghorn samples (76%) than in the broiler samples (62%). The used identification and selection method worked very well compared with the tools and results reported by other authors. Picoult-Newberg et al. (1999) analyzed 18 human individuals for the presence of 88 SNP markers and validated 55 of these (63%). Buetow et al. (1999) analyzed 10 human individuals for the presence of 192 candidate SNP and validated 82%. Batley et al. (2003) analyzed 4 inbred corn lines to analyze the existence of 264 SNP, and were able to validate 91% of the SNP. Therefore the identification and selection method reported in the current study has resulted in a collection of 5,332 reliable SNP. In addition, the 5,332 chicken SNP were selected based on having a high frequency of both alleles in the population and will be well suited for the construction of a high-density chicken SNP map.

ACKNOWLEDGMENTS

The current research was funded by a joint grant from Wageningen University and Research Centre and the Faculty of Veterinary Medicine of the Utrecht University.

The authors thank M. A. M. Groenen, Wageningen University and Research Centre, The Netherlands, for stimulating discussions and for providing the genomic samples used in this study.

REFERENCES

Altschul, S. F., W. Gish, W. Miller, E. W. Meyers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Andersson, L., and M. Georges. 2004. Domestic-animal genomics: Deciphering the genetics of complex traits. Nat. Genet. 5:202–212.

Batley, J., G. Barker, H. O'Sullivan, K. J. Edwards, and D. Edwards. 2003. Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. Plant Physiol. 132:84–91.

Boardman, P. E., J. Sanz-Ezquerro, I. M. Overton, D. W. Burt, E. Bosch, W. T. Fong, C. Tickle, W. R. A. Brown, S. A. Wilson, and S. J. Hubbard. 2002. A comprehensive collection of chicken cDNAs. Curr. Biol. 12:1965–1969.

Buetow, K. H., M. N. Edmonson, and A. B. Cassidy. 1999. Reliable identification of large numbers of candidate SNPs from public EST data. Nat. Genet. 21:323–325.

Emara, M. G., and H. Kim. 2003. Genetic markers and their application in poultry breeding. Poult. Sci. 82:952–957.

Ewing, B., L. Hillier, M. C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8:175–185.

Gordon, D., C. Abajian, and P. Green. 1998. Consed: A graphical tool for sequence finishing. Genome Res. 8:195–202.

Groenen, M. A. M., H. H. Cheng, N. Bumstead, B. F. Benkel, W. E. Briles, T. Burke, D. W. Burt, L. B. Crittenden, J. Dodgson, J. Hillel, S. Lamont, A. Ponce de Leon, M. Soller, H. Takahashi, and A. Vignal. 2000. A consensus linkage map of the chicken genome. Genome Res. 10:137–147.

Huang, X., and A. Madan. 1999. Cap3: A DNA sequence assembly program. Genome Res. 9:868–877.

Kim, H., C. J. Schmidt, K. S. Decker, and M. G. Emara. 2003. A double-screening method to identify reliable candidate nonsynonymous SNPs from chicken EST data. Anim. Genet. 34:249–254.

Lamont, S. J. 2003. Unique population designs used to address molecular genetics questions in poultry. Poult. Sci. 82:882–884.

Malek, M., and S. J. Lamont. 2003. Association of *INOS*, *TRAIL*, *TGF-β2*, *TGF-β3*, and *IgL* genes with response to *Salmonella enteritidis* in poultry. Genet. Sel. Evol. 35:S99–S111.

Marth, G. T., I. Korf, M. D. Yandell, R. T. Yeh, Z. Gu, H. Zakeri, N. O. Stitziel, L. Hillier, P.-Y. Kwok, and W. R. Gish. 1999. A general approach to single-nucleotide polymorphism discovery. Nat. Genet. 23:452–456.

Meldrum, D. 2000. Automation for genomics, part two: Sequencers, microarrays, and future trends. Genome Res. 10:1288–1303.

Nickerson, D. A., V. O. Tobe, and S. L. Taylor. 1997. Polyphred: Automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res. 25:2745–2751.

Oliphant, A., D. L. Barker, J. R. Struelpnagel, and M. S. Chee. 2002. Beadarray technology: Enabling an accurate, cost-effective approach to high-throughput genotyping. Biotechniques 32:S56–S61.

Pertea, G., X. Huang, F. Liang, V. Antonescu, R. Sultana, S. Karamycheva, Y. Lee, J. White, F. Cheung, B. Parviz, J. Tsai, and J. Quackenbush. 2003. TIGR gene indices clustering tools (TGICL): A software system for fast clustering of large EST datasets. Bioinformatics 19:651–652.

Picoult-Newberg, L., T. E. Ideker, M. G. Pohl, S. L. Taylor, M. A. Donaldson, D. A. Nickerson, and M. Boyce-Jacino. 1999. Mining SNPs from EST databases. Genome Res. 9:167–174.

Ren, C., M.-K. Lee, B. Yan, K. Ding, B. Cox, M. N. Romanov, J. A. Price, J. B. Dodgson, and H.-B. Zhang. 2003. A BAC-based physical map of the chicken genome. Genome Res. 13:2754–2758.

Rozen, S., and H. J. Skaletsky. 2000. Primer3 on the WWW for general users and biologist programmers. Methods Mol. Biol. 132:365–386.

Staden, R., K. F. Beal, and J. K. Bonfield. 1998. The Staden Package. Pages 115–130 in Computer Methods in Molecular Biology. S. Misener, and S. A. Krawetz, ed. The Humana Press Inc., Totowa, NJ.

Tang, K., D.-J. Fu, D. Julien, A. Braun, C. R. Cantor, and H. Köster. 1999. Chip-based genotyping by mass spectrometry. Proc. Natl. Acad. Sci. USA 96:10016–10020.

Vieux, E. F., P.-Y. Kwok, and R. D. Miller. 2002. Primer design for PCR and sequencing in high-throughput analysis of SNP's. Biotechniques 32:S28–S32.

Vignal, A., D. Milan, M. SanCristobal, and A. Eggen. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. 34:275–305.

Wang, D. G., J.-B. Fan, C.-J. Siao, A. Berno, P. Young, R. Sapolsky, G. Ghandour, N. Perkins, E. Winchester, J. Spencer, L. Kruglyak, L. Stein, L. Hsie, T. Topaloglou, E. Hubbell, E. Robinson, M. Mittmann, M. S. Morris, N. Shen, D. Kilburn, J. Rioux, C. Nusbaum, S. Rozen, T. J. Hudson, R. Lipshutz, M. Chee, and E. S. Lander. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077–1082.