4 Results

Physical maps provide a view of the genome in terms of ordered genomic clones. A complete map consists of a series of continuously overlapping cloned genomic DNA segments. Previously cloned genes and DNA markers on a genetic map can be localized to a specific clone on a physical map by hybridization techniques. This provides a way to estimate locations of different markers and distances between them. This information simplifies the process of positional cloning. For instance, if one gene is linked to two DNA markers and the physical map between these two markers is complete, then clones in this area of the physical map must contain this gene. Additionally, physical maps also simplify the process of entire genome sequencing using the clone by clone approach.

This project was initiated to create a physical map of the medaka genome with overlapping BAC (bacterial artificial chromosome) clones, covering all medaka chromosomes. To find overlapping clones, a hybridization strategy was employed. The utilized probes were derived from three different sources: BAC end-fragments of clones of medaka genomic libraries, medaka cDNA fragments and 35-mer oligonucleotides, specific for medaka ESTs aligned against fugu genome scaffolds.

4.1 Creation of BAC end-fragments

The first source of probes to create the medaka physical map consisted of probes derived from medaka BAC end-fragments. The inserts of BAC clones were flanked by two known sequences (primers), which were used to amplify these fragments. These probes were generated as described below, sequenced and have been used as markers to construct a physical map.

4.1.1 BAC DNA isolation

The BAC system is an important tool for molecular genetic research. Although the YAC system can hold large-insert DNAs (1000kb), using the BAC system was preferred for our approach as BAC clones are amplified in bacterial cells and isolated and manipulated simply with basic bacterial plasmid technology and because they also show lower levels of chimerism compared to YAC clones (Matsuda *et al.*, 2000).

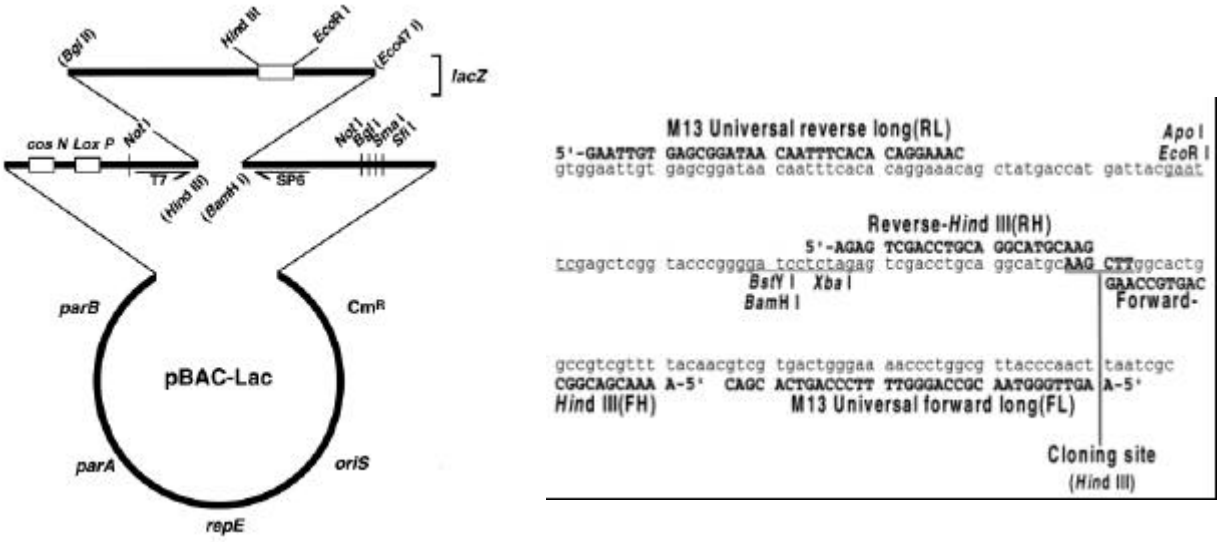

Figure4-1: (Left) The map of pBAC-Lac vector (Asakawa *et al.*, 1997). (Right) Nucleotide sequences around cloning site of pBAC-Lac.

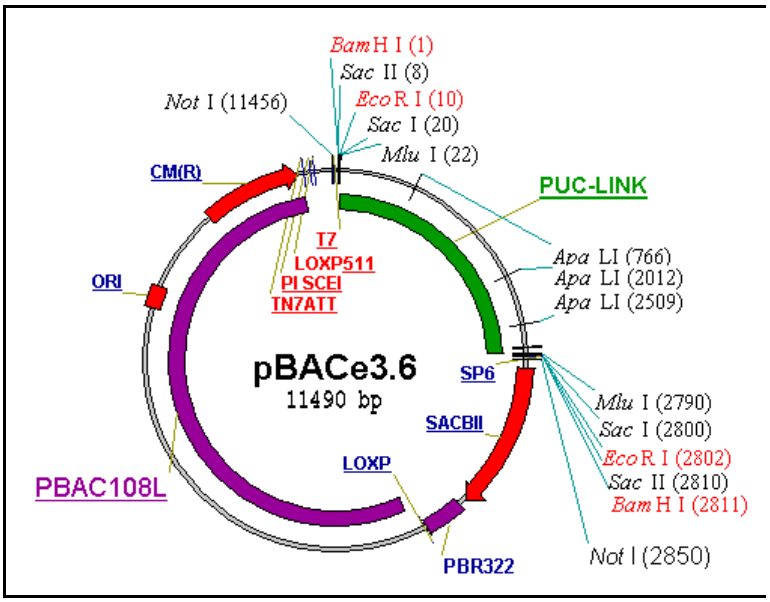

Figure 4-2: The map of pBACe3.6

To construct the medaka physical map altogether three different medaka BAC libraries were utilized, the Cab and Hd-rR libraries both created from a strain of the southern population and the HNI library from a northern strain (see 2.2.2). The Hd-rR (Matsuda *et al.*, 2000) and HNI (Kondo *et al.*, 2002) libraries had been cloned into pBAC-Lac (Asakawa *et al.*, 1997) vectors (Figure4-1) and the Cab library (See 2.4.6.1) in pBACe3.6 (Figure 4-2). The BAC vectors all confer resistance to chloramphenicol.

The plasmid DNA of BAC clones was isolated as described in 3.10. BAC is a system for isolating cloning genomic DNA based on the F-factor plasmid, the bacterial sex or fertility plasmid. Since the F factor is generally maintained at only 1-2 copies per cell, the BAC DNA yield obtained from a given volume of culture is much lower than that for other plasmid DNA. Therefore to save the isolated DNA for further experimental steps like digestion, ligation and amplification, the quality and quantity of DNA isolation was not checked here.
Isolated BAC DNA was digested with the restriction enzyme RsaI. RsaI was chosen because of its 4 bp recognition site which occurs quite often in the DNA sequence (average size of RsaI-fragments is about 250 bp) and also produces the DNA fragments with blunt ends. Digested BAC DNA inserts were cloned into the multiple cloning site (MCS) of the pBluescript II KS vector in the position of the EcoRV restriction site.

4.2 Amplification

The BAC end-fragments were created by polymerase chain reaction (PCR) using two primers, one primer specific for the BAC vector and another primer matching pBluescript (Figure 4-3). Amplifications were carried out for both ends of DNA inserts as described in 3.16.1.
It was observed that amplification of some end-fragments using the routine primers did not produce specific bands. In this case, nested primers (3.6.1) were designed and after amplifying, specific bands were created.
With these methods the DNA of 20 plates (384 MTP) of the Cab library (plate 1-20), 11 plates (384 MTP) of the Hd-rR library (145-155) and 7 plates of the HNI library (plate 73-79) was isolated, digested, ligated and both end–fragments were amplified. The amplified fragments have variable sizes but most of them were in the range of 250 –1300 bp. Figure 4-4 shows a 0.7% agarose gel of BAC end-fragments. As standard marker, 0.5 µg of ?X174 DNA/BsuRI was used.

The amount of amplified BAC end-fragments from each plate was variable. From some plates up to 87% of clones were amplified at both ends. The quality of amplified fragments was checked on an agarose gel. Specific fragments, which could be amplified from both ends and that were of sizes larger than 200 bp were selected and purified using two different kits, Silica Matrix DNA-Purification and NC001 nuclean I kits. In the case of using the QIAquick kit, the loss of DNA was less than with the NC001 nuclean I kit. As a consequence, the use of QIAquick kit was preferred to the other one. The purified BAC end-fragments were sent to the lab of Prof. Shimizu, Keio University, Japan, for sequencing.

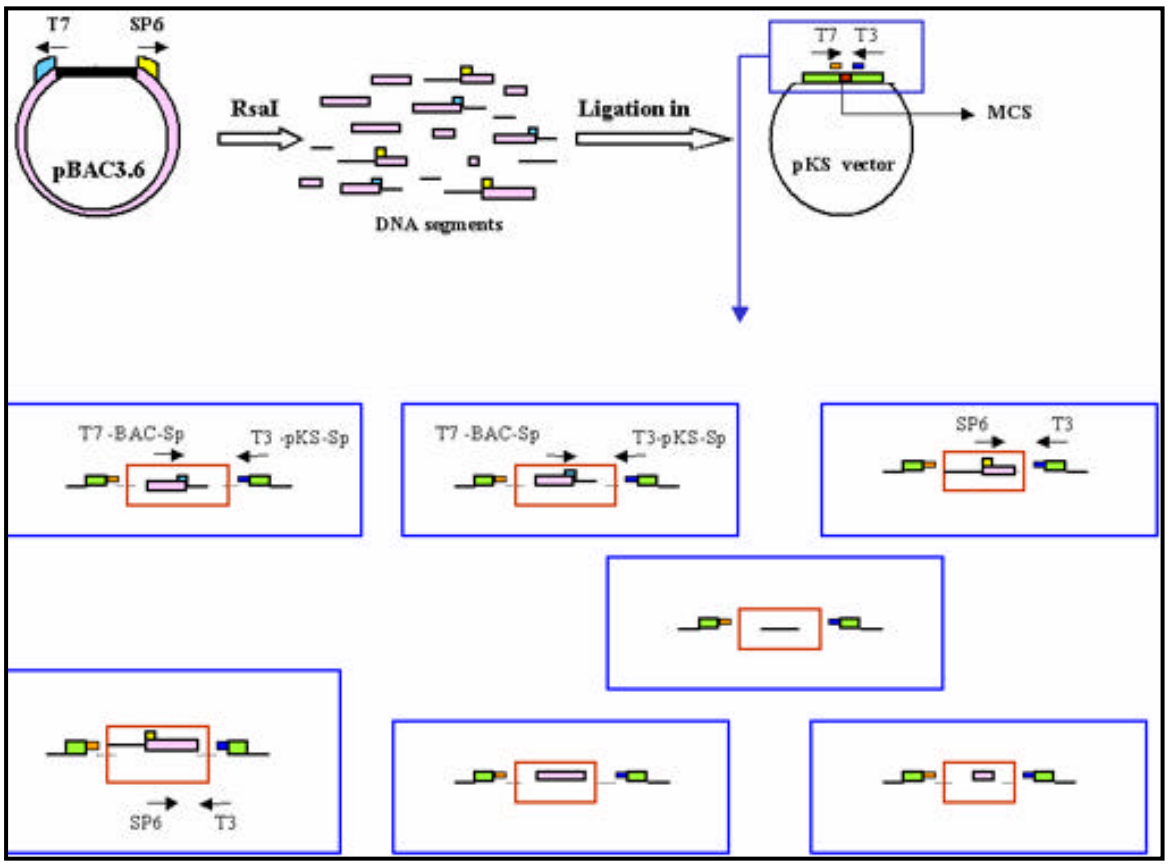

Figure 4-3: Generation of BAC end-fragments: The isolated BAC DNAs were digested with RsaI. RsaI cuts the DNA many times and DNA segments of different size and different composition result. The DNA segments were ligated into the pKS vector Ligated fragments were amplified using one primer specific for the BAC vector and another primer matching to pBluescript. The possibilities of different kinds of amplification are illustrated here. Only fragments, which bind one of the BAC primers in addition to a Bluescript primer, can be amplified.

During this time the genome sequence of the Japanese pufferfish, *Takifugu rubripes*, which is closely related to the medaka, was published (Aparicio *et al.*, 2002) and for further work the synteny of these two fish genomes was used to create another group of probes as physical markers (See 4.5). Production of oligonucleotides based on expressed sequences was easier and quicker than the production of BAC end-fragments and the generation of these probes was stopped.

4.3 Hybridization

4.3.1 Hybridization with DIG labeled probes

To hybridize the BAC end-fragments against BAC libraries, both non-radioactive and radioactive approaches were tested. First a non-radioactive approach with digoxigenin (DIG-dUTP) was employed, because it requires less safety precautions than applying the radioactive procedure. The Cab end fragments were labeled with DIG-dUTP by PCR as described in (3.24.1).

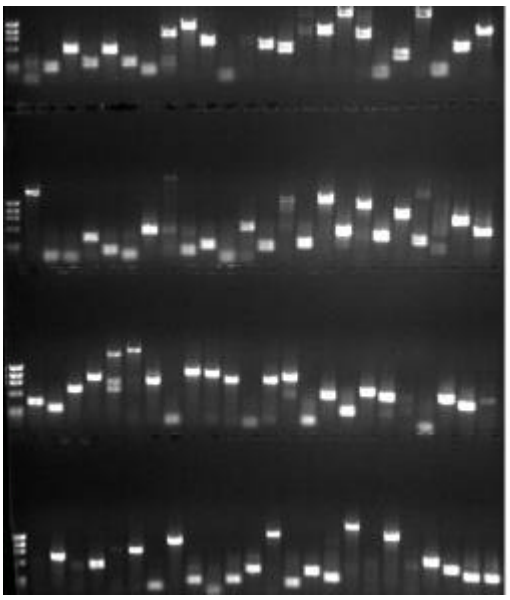

Figure 4-4: BAC end-fragments from CAB library plate 11-1 amplified using SP6-long and T3-long primers.

The PCR products were checked on an agarose gel and it was observed that several unspecific products of different sizes were obtained. To solve this problem, new PCR primers (nested SP6 and nested T3) were designed and the PCR was repeated. Nested PCR was successful and only one specific band for each sample was created. Afterwards the PCR products were hybridized against Cab filters, which were spotted by the RZPD (Resource Centre/Primary Database) but no signal was detected. The experiment was repeated under different conditions, changing the time of incubation and concentration of buffer and solutions, but no reasonable results were observed. To compare both non-radioactive and radioactive methods, the same probes were labeled radioactively (4.3.2) and hybridized against Cab filter. Hybridizations were successful and positive signals were detected in the radioactive approach, which was therefore employed for further work.

4.3.2 Hybridization utilizing the radioactive approach

For hybridization of BAC end-fragments against BAC library filters, the Feinberg and Vogelstein method (3.24.2.1) was used employing ? -^{32}P-dCTP isotope. During first experiments, strong backgrounds prevented filter analysis. Background may result because of either bad quality of probes, materials or chemicals, which were used for this experiment. To remove the backgrounds two actions were taken, firstly purification of PCR products to improve the quality of probes, using the QIAquick PCR Purification Kit before labeling. Secondly, the performance of the random priming kit for labeling the probes was compared to the traditional Feinberg and Vogelstein method. All possible four combinations (PCR purification or no PCR purification combined with use of kit or Vogelstein method) were tested by hybridizing the same sample to four different Cab filters. As result, it was observed that the hybridization of the probes generated by using the kit and purified after PCR, was better than the others.

Although the use of a kit and purified probes improved the results of hybridizations, the background was not totally removed. Other possibilities for appearance of background are the existence of repetitive elements or vector fragments in template DNAs amplified by PCR and consequently labeled and hybridized to all or several clones. Therefore possible similarities between pBluescript from which PCR inserts were amplified and pBACe3.6 and pBAC-lac, containing BAC inserts, were checked via the blast program (NCBI), but no homology was found. It is known that by using competition DNA it is possible to mask repetitive elements or remaining vector fragments, therefore labeled DNA can bind more specifically to BAC inserts. As competition DNA, genomic DNA of medaka (3.12) and vector DNA (pBACe3.6 or pBAC-Lac vector, without insert or with small insert) were used. By the means of this method (3.24.2.9) the background was mostly removed.

4.4 Creating cDNA fragments

The second part of resources, which were used to establish a medaka physical map, came from medaka cDNA clones. 445 medaka cDNA clones were obtained from the group of Hiroshi Mitani, University of Tokyo, Japan. These clones had been genetically mapped and were used as anchors for our map. They were derived from four libraries, which had been prepared from whole body, liver, ovary and a cultured cell line of the HNI strain (Table 3-1). The cDNAs had been cloned in three different vectors: pME18s (Ola clones), pBluescript SK(+)(Olb and Olf clones) and pUC118 vector (Olc, Old and Ole clones) (see

http://mbase.bioweb.ne.jp/~dclust/lib_info.htm). Inserts were amplified by PCR with respective primers (3.6.2) and PCR products were checked on a 0.7% agarose gel (Figure 4-5). 330 clones, which after amplification showed only one band, were chosen and purified using NC001 nuclean I kit. Then the purified PCR products were labeled and hybridized against the Cab library filters. To establish the hybridization approach, the same experiments for testing hybridization quality were done as in the case of the BAC end-fragments. Finally we chose to use purified PCR products and random primed kit for labeling and competition with genomic fish DNA and vector DNA to reduce the background.

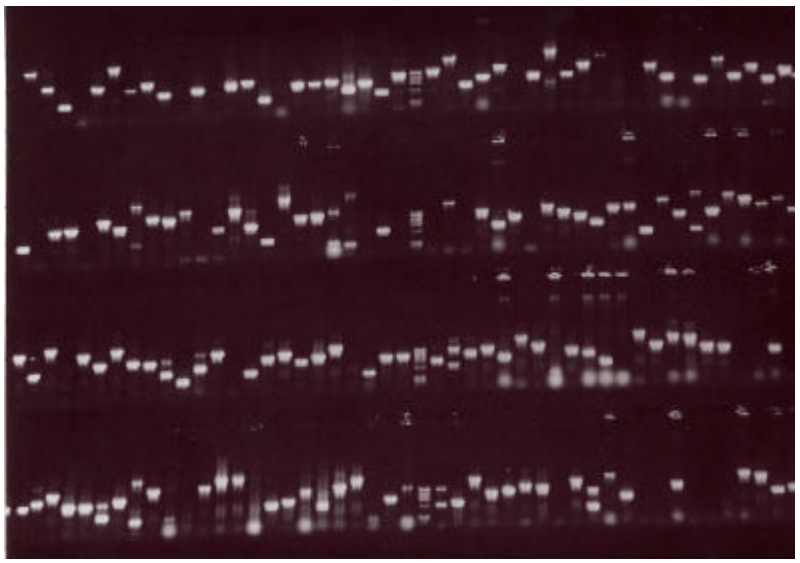

Figure 4-5: cDNA-fragments from plates 3 and 4 obtained from Hiroshi Mitani.

Of these 59 fragments did not hybridize against any BAC clones, probably because of the bad quality of the filters. Anyway, we stopped to amplify and hybridize the cDNA fragments, because pooling of the probes was not possible using the probes one at the time did not provide the required high throughput.

Total number of cDNA clones obtained	Clones with single band after PCR	Successfully hybridized (at least one BAC positive)
445	330 (74%)	271 (61%)

Table 4-1: Percentage of clones used in experiment compared to total number of cDNA clones.

Of 445 medaka cDNA clones, which were amplified by using of appropriate primers, 330 had only one band. 115 clones were either empty (without insert) or showed unspecific amplification results. The 330 cDNA fragments were chosen to hybridize against Cab BAC-library filters (Table 4-1).

4.5 Use of the synteny between medaka and *Takifugu rubripes* sequence to identify probes for mapping

The third and main source of probes, which were utilized to generate the medaka physical map, was derived from a comparison of synteny in genomes of medaka and *Takifugu rubripes.* Fugu is one of the first vertebrates and the first fish that had its genome completely sequenced (Aparicio *et al.*, 2002) (see 2.5.3).

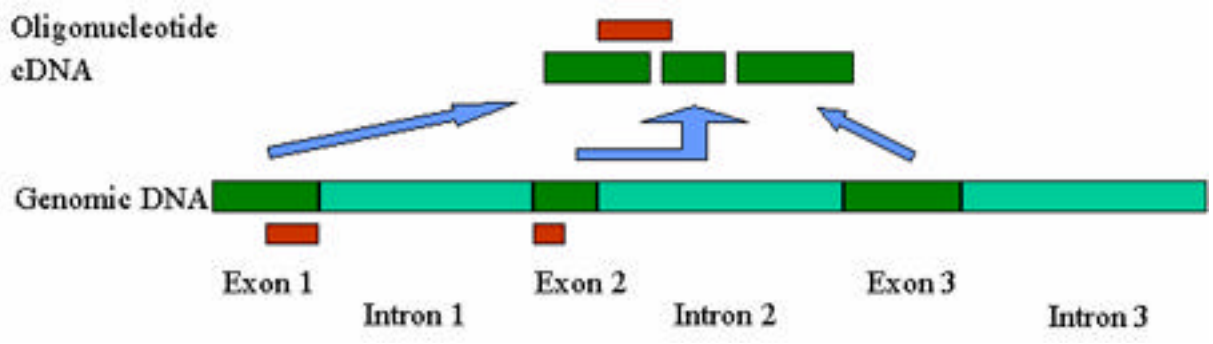

Figure 4-6: If oligonucleotide probes are generated from cDNA sequences, exon boundaries have to be taken into account. Exon structures are well conserved between vertebrate species. A cDNA sequence from one species can therefore be aligned to the genome from another species to information on the exonic structure of a particular gene. Once the exonic structure is known, an oligonucleotide can be designed that located within one exon.

The fugu genome v3.0 published in August 2002 consisted of 20, 379 sequence scaffolds, which covered 329.7 Mb of the fugu genome. Figure 2-1 shows that medaka and *Takifugu rubripes* are closely related and that their estimated evolutionary distance is 60-80 Myr. Considering that genes and gene order on chromosomes remain largely conserved during the evolution between species, synteny between medaka and fugu genome should be highly probable. Consequently, the fugu genome can be used as a template for constructing the physical map of medaka. The medaka cDNA sequences were aligned against the fugu genomic sequence to permit the design of an oligonucleotide located in one single exon, because oligonucleotiedes designed over exon-exon borders will not hybridize on genomic DNA filters (Figure 4-6).

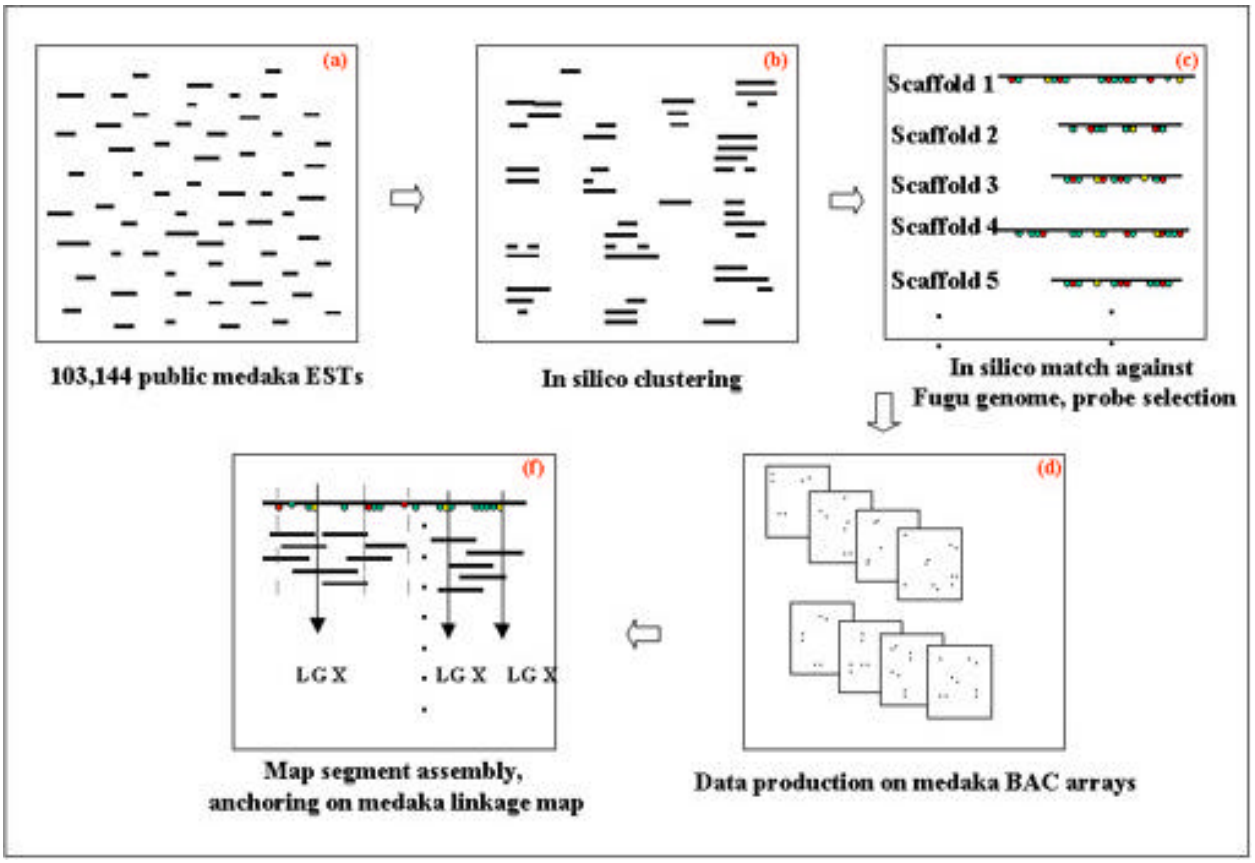

Figure 4-7: The strategy to create a physical map for the medaka genome utilizing fugu-medaka synteny. (a, b) The medakas ESTs, available at NCBI, were clustered in silico. (c) They were aligned against fugu scaffolds Red and yellow circles are EST clusters from which probes are generated. For yellow clusters, genetic mapping information is available. From these probes 35-mer oligos were designed. (d) The probes were hybridized against different BAC libraries to create a data set. (f) BAC contig construction. In the present example, one fugu scaffold translates into two medaka contigs, both of which are linked to same medaka chromosome by genetic mapping.

Furthermore, as the introns have been removed in cDNA, oligonucleotides can not be designed directly on cDNA sequences. As the genes are not equally distributed along the genome, a comparison between medaka and fugu allows the synthesis of oligonucleotides on the genome in an even way, covering the medaka genome approximately every 50 –100. kb.

At the time of publication of the fugu genome, a large number of medaka ESTs had been deposited at the NCBI database. 103,144 medaka EST sequences were downloaded from this website. The strategy applied to create the medaka physical map is summarized in Figure 4-7.

4.5.1 Medaka EST data clustering and match against fugu genome sequence

The 103,144 medaka ESTs were grouped into 21,121 pre-clusters based on all-against-all comparison by blastn (3.25.1). Then, these pre-clusters were grouped to final sequence clusters, and consensus sequences were calculated by the cap3 (Huang and Madan, 1999) program. Cap3 found 9379 true sequence clusters with more than one EST. 11,742 EST sequences remained as singletons not grouped into any clusters.

To use the synteny between fugu and medaka to create the third group of probes, the 21,121 medaka unique sequences were aligned against the fugu genome draft using crossmatch (http://www.genome.washington.edu/UWGC/analysistools/Swat.cfm).

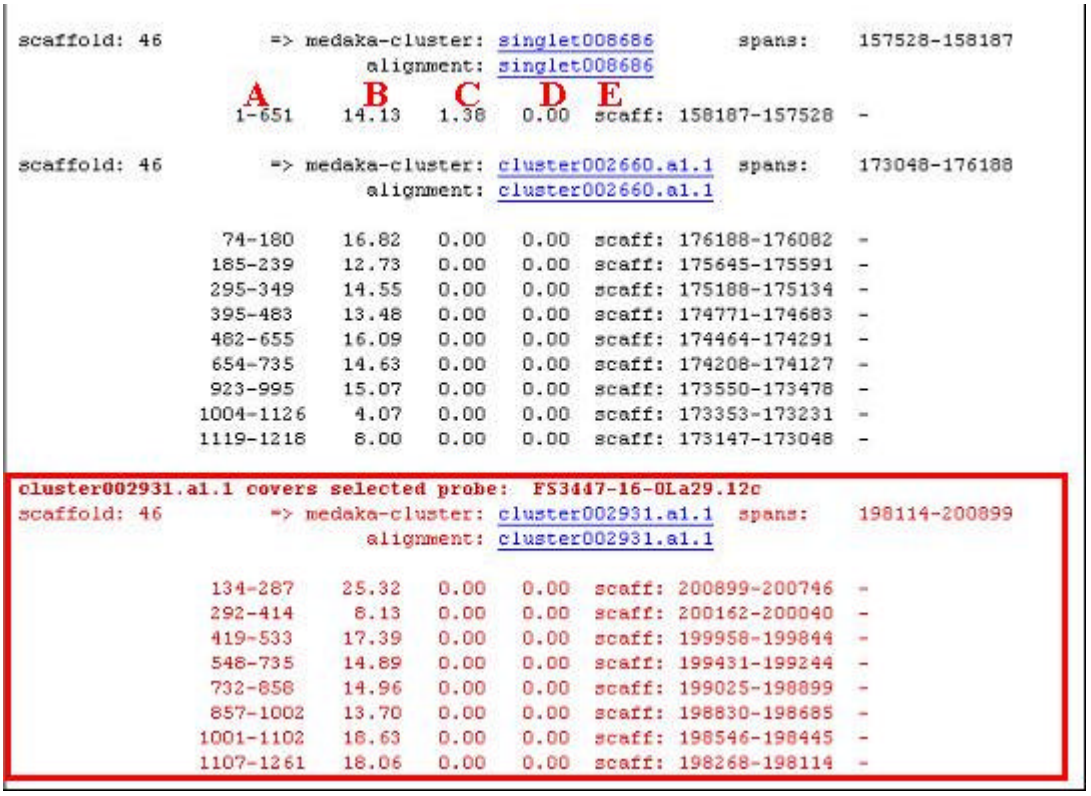

Figure 4-8: Medaka clusters aligned against the fugu genome using cross_match. See the text. Columns: A, match region in Medaka-cluster (MC). B, percentage of substitutions in MC relative to scaffold. C, percentage of deletions. D, percentage of insertions. E, match region in scaffold. The "-" in the last column indicates that the match is on the reverse strand.

In the early stage of the project, up to August 2002, the fugu v2.0 assembly was used for the alignment. The majority of probes were designed by aligning medaka EST clusters against the

v3.0 assembly of the fugu genome sequence. The comparison of medaka sequences with fugu genome scaffolds is available at (http://www.molgen.mpg.de/~hennig/medak/fugu-v3-update-12-2003/SCAFF-HITS-table.html). Specifically, fugu scaffold numbers in assembly v2.0 and v3.0 are unrelated. Table 3-1 shows an example of two medaka clusters and one singleton aligned to fugu scaffold 46. Sequences and information about medaka clusters and fugu scaffolds are available by double clicking on them. The probe names contain information about the respective syntenic position in fugu sequence scaffolds.

For example for the probe FS3447-16-OLa29.12c, which is marked in red box (Figure 4-8):

FS3447: v2.0 assembly, scaffold number 3447

16: kb position 16 in fugu scaffold 3447

OLa29.12c: cluster identifier, refers to MBase-marker OLa29.12c

11,254 of medaka clusters or singletons, which corresponded to 53% of all unique sequences, successfully matched a fugu scaffold. 3397 fugu scaffolds, which correspond to 249 Mb of fugu sequence or 75% of the total assembly, were hit by at least one medaka cluster.

4.5.2 Selection of 35-mer oligonucleotides, pooling and hybridization

For many of medaka genes, EST sequences are available but their position within the genome is not clear. To convert EST sequences to markers, a short sequence representative to a certain EST cluster was designed as a probe and hybridized against colony filters of BAC libraries. Positive hybridization signals announced the location of the gene in the BAC clones of the library.

To find the optimal length of oligonucleotides, five sets of oligonucleotides with different length of 25, 30 and 35 nucleotides for each EST, were designed. Each oligonucleotide was separately hybridized against BAC libraries. The hybridization was done also in pooling format that means all five oligos from each 25, 30 and 35mer were pooled and used as probe for hybridization. To check that all signals from the hybridization with single probes were also represented in the image of their pools, the related autoradiograms were compared against each other. The image of 25mer pool showed missing of a lot of signals and the background was moreover strong. Although the background was not observed in the image of pool of 30mer oligos, a lot of signals were missing in this image. The best results were observed in the pool of 35mer oligos, which represented all the signals from each single hybridization (Figure 4-9).

Then for further experiments 35mer oligonucleotides from genes were designed for hybridization using the alignment of medaka cDNA sequences against the fugu genome

(http://www.molgen.mpg.de/~hennig/medak/fugu-v3-update-12-2003/SCAFF-HITS-table.html).

1072 fugu scaffolds from number 1-1500 were employed to design 2086 35-mer probes with at least one probe per scaffold. Moreover, 191 further 35-mer probes from the fugu scaffold with higher number than 1500 were also designed. Most of these 191 probes corresponded to genetically mapped medaka genes according to MBase. The lists of 35-mer oligonucleotides are available at the PPE (Pooled Probe Editor) website (http://goblet.molgen.mpg.de/cgi-bin/hb/hb2004.cgi).

Altogether 2363 35-mer oligonucleotides were designed (3.22). To reduce the number of hybridizations and to accelerate the work the 35-mer probes were pooled according to a 2D and 3D scheme (Figure 3-2 and Figure 3-3). At the beginning of work the probes were pooled according to 2D. The 2D pooling scheme consisted of row pool (RP) and column pool (SP). Each RP consisted of six 35-mer probes and each SP of eight probes (Figure 4-10). 25 pmol of each pooled probe was labeled by ?-^{32}P-dATP, using T4-polynucleotide kinase (PNK) (see 3.24.2.3). Later, we increased the number of probes in each pool using a 3D scheme, which consisted of block row pool (BRP), block column pool (BSP) and block plate pool (BPP) (see Figure 3-3). Consequently each BRP consisted of 36, each BSP of 48 and each BPP of 48 probes. Because of the large number of oligonucleotides in each probe pool when using 3D pooling scheme, the use of ^{33}P (? -^{33}P-dATP) isotope is preferred to ^{32}P for labeling reactions because of its lower energy and therefore lower radiation range. This makes it possible to evaluate weak signals even when they are close to strong signals. Most importantly, the use of ^{33}P-labeled probes reduces blurring effects and clone coordinates can be determined automatically to a large extend using appropriate image analysis software.

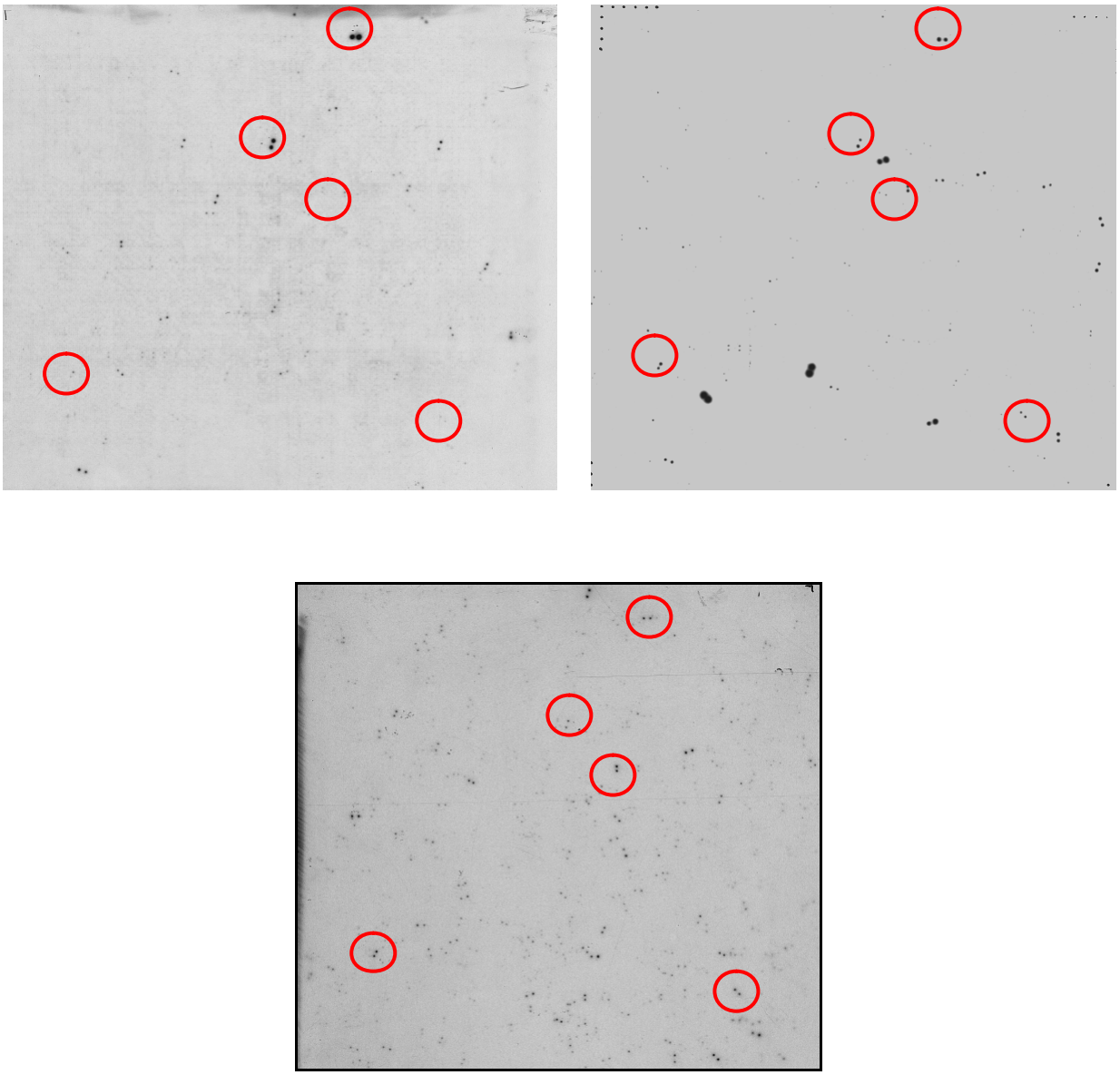

Figure 4-9: Autoradiograms of pools of the same oligonucleotides with different lengths. Above left: 25mer, above right: 30mer and below: 35mer. Some common signals in all three autoradiogrames are marked by red circles.

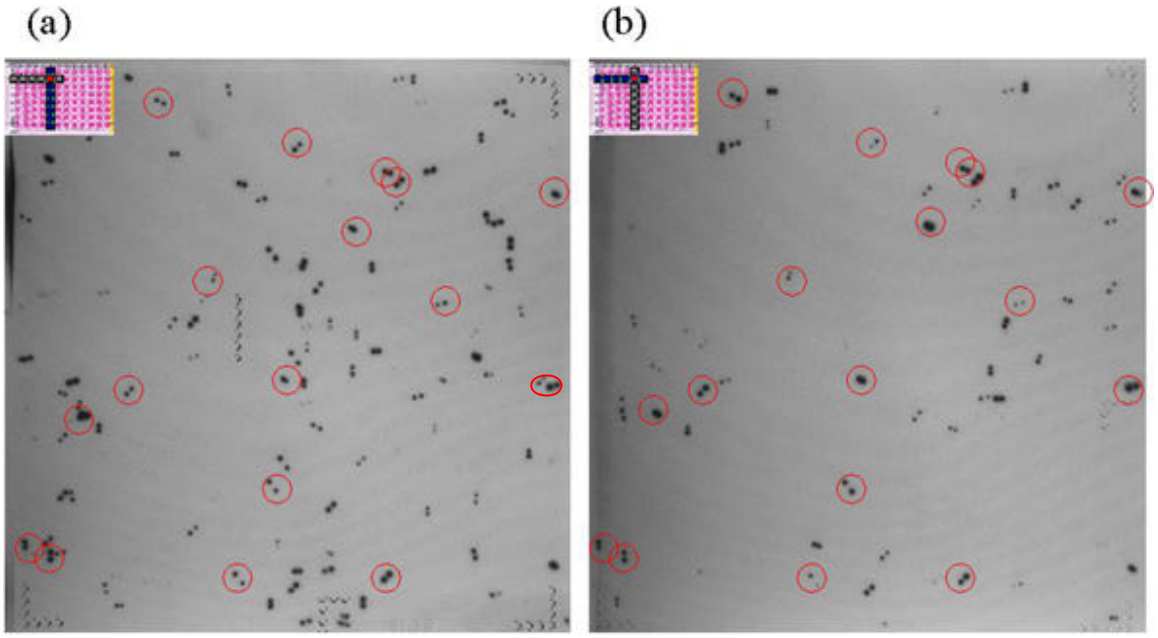

Figure 4-10: Autoradiograms created by hybridization of probes according to a 2D pooling scheme. a: Hybridization of column pool (SP), which consisted of 8 probes (96-well plate position A5-H5). b: hybridization of row pool (RP), which consisted of 6 probes (96-well plate position B1-B6). The films obtained from RP and SP compared against each other. Selected signals (red circles) belong to the probe (FS449-23-Olc58.12d on LG21, map segment 16) in the microtiterplate.

4.5.2.1 Colony filters

Colony filters were spotted and processed as described in 3.21. In this way more than 750 filters were spotted with Cab, Hd-rR and HNI BAC clone libraries. 27,648 BAC clones from Cab strain were spotted on the Cab filters using a 5x5 pattern (Figure 3-1). The average insert size of BAC clones from Cab library is 150 kb. Assuming a size of 800 Mb for the medaka genome, it was predicted that Cab BAC clones cover the medaka genome 5-fold.

The Hd-rR filters were spotted with 36,864 BAC clones (6x6 spotting pattern). With an average insert size of 210 kb, these clones cover the medaka genome 9-fold. Cab and Hd-rR libraries together cover therefore the medaka genome approximately 14-fold.

Our estimation was confirmed by results of the hybridization of 2363 35-mer oligonucleotides against Cab and Hd-rR libraries. Figure 4-11 illustrates the number of hybridized clones per probe from these two libraries. Close to what is expected, the average number of hybridized Cab clones per probe is 4.8 and for Hd-rR is 10.5. In total, 41,882 BAC clones (65% of clones used, assuming no empty plate positions) were hit by at least one probe. 86 of 2363 35-mer

oligonucleotides (3.6%) did not hybridize to any clone in the libraries. On the contrary, 153 probes hybridized with more than 30 clones.

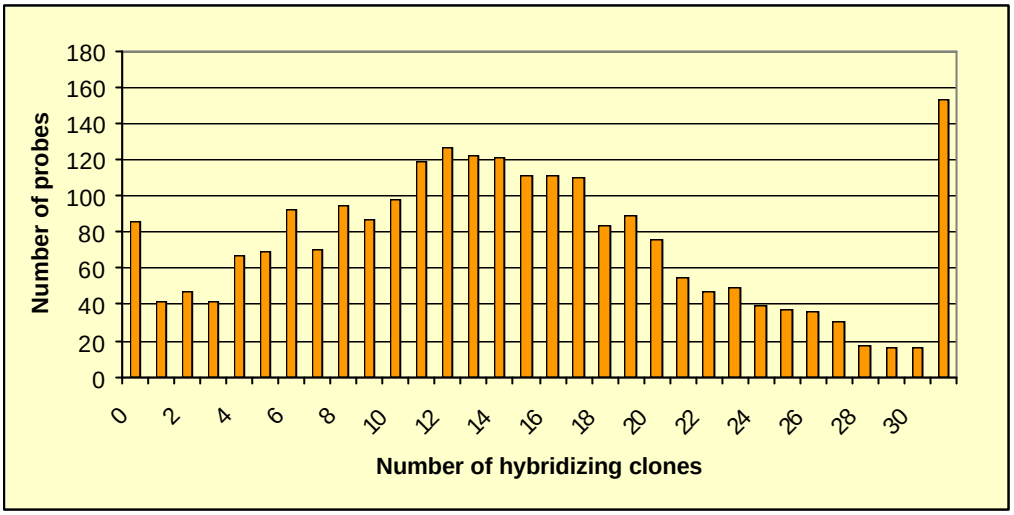

Figure 4-11:Number of hybridizing clones per probe for both Cab and Hd-rR libraries. These two resources represent together a 14-fold coverage of the medaka genome. 86 clones did not hybridize with any clones and 153 probes hybridized with more than 30 clones.

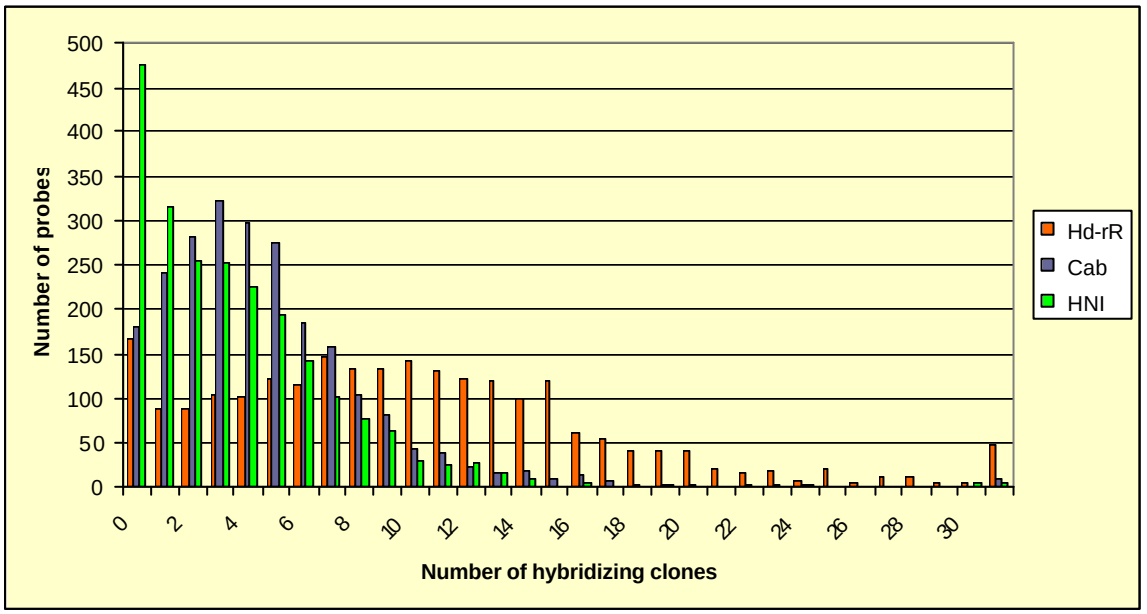

Figure 4-12: The number of hybridizing clones per probe detected in the three libraries Cab, Hd-rR and HNI.

On the HNI filter 27, 648 HNI BAC clones were spotted using a 5x5 spotting pattern. The average insert size of these clones is 160 kb, predict to cover the medaka genome five times. 2304 35mer oligonucleotides were hybridized against HNI filters according to the 3D scheme.

The hybridization of probes against the three different BAC libraries, Cab, Hd-rR and HNI were compared against each other (Figure 4-12). This figure shows that the number of hybridizing clones per probe was less in the HNI library than in the other libraries and it is due to use of the automatical image analysis software for the first time, which could not detected the week signals. For the further experiments this problem was resolved.

4.5.3 Data analysis

The obtained autoradiograms were scanned and image files were saved as tiff format. The image files were imported into the program VisualGrid. Via this software, grids with a spotting format for Cab, Hd-rR and HNI were manually laid onto the image files. For further image analysis, two different software tools were employed.

No	Short_Name	BlockX	BlockY	PlateNo	PatternNo
1	22P16	0	22	4	3
2	70P3	0	70	12	3
3	26N3	2	26	5	3
4	34K18	5	34	6	3
5	62J20	6	62	11	3
6	34H7	8	34	6	3
7	40G12	9	40	7	3
8	70B18	14	70	12	3
9	38B2	14	38	7	3
10	32B1	14	32	6	3
11	6P11	16	6	1	3
12	6608	17	66	11	3
13	12H1	19	12	2	3
14	43L11	20	43	8	3
15	43J16	22	43	8	3
16	66J16	22	66	11	3
17	36J12	22	36	6	3
18	67H5	24	67	12	3
19	13F23	26	13	3	3
20	55E15	27	55	10	3

Figure 4-13: RST file from a Cab filter hybridized with the RP1 probe of Cab library plate 15. The coordinates of each signal in microtiter plate were calculated by VisualGrid, shown in the red box.

The films, created by hybridization of probes according to the 2D pooling scheme, were subsequently analyzed by VisualGrid. Positive signals were identified and selected by clicking. Selected positives were exported into a textfile (TXT) and saved again as Non-standard Residue (RST) file (Figure 4-13).

VisualGrid calculated the coordinates of clones in microtiterplates (Figure 4-13). The position of the detected clone in the microtiter plate is shown in the second column marked in red box. For instance, the first signal in Figure 4-13 belongs to the Cab plate number 22 in the position P16.

On the films, which were obtained from probe pools of the 3D pooling scheme, there were lots of signals, which were difficult to be analyzed manually. These image files where first imported into VisualGrid, overlaid manually with a grid and were further automatically processed (3.25.2) using in-house developed tools (Steinfath *et al.*, 2001), after which clone coordinates were calculated and exported as RST files.

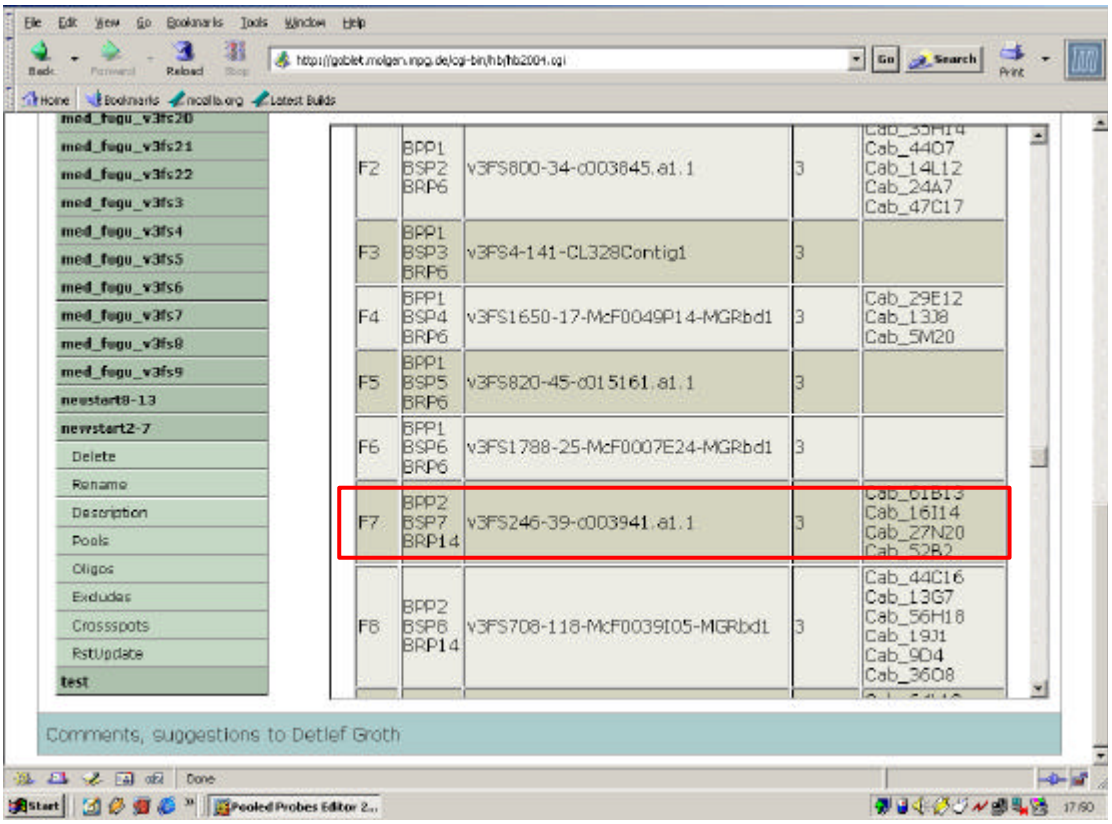

Figure 4-14: Pooled Probe Editor tool: Web interface to inspect PPE results. The RST files of BSP, BRP and BPP were compared against each other and the signals in common were found, corresponding to the plate position of intersection of these three pools.

These data were imported into the Pooled Probe Editor (PPE) software (Figure 4-14). PPE (http://goblet.molgen.mpg.de/cgi-bin/hb/hb2004.cgi) compared the hybridization data (RST

files) of RP against SP or BSP, BRP and BPP against each other and identified the clones, which these pools had in common.

For example Figure 4-15 shows the hybridization of BPP2, BRP14 and BSP7 against three Cab colony macroarrays. The clone lists, in RST file format generated by the image analysis software, were compared against each other by PPE and the clones, which were in common in three pools, were identified. For instance, Figure 4-15 shows clones, which are shared between 3 Cab filters hybridised with pools BPP2, BRP14 and BSP7. The intersection between these three pools is marker v3FS131-283-McF0035O07-MGRbd1, in the position 1F7 in microtiterplate1.

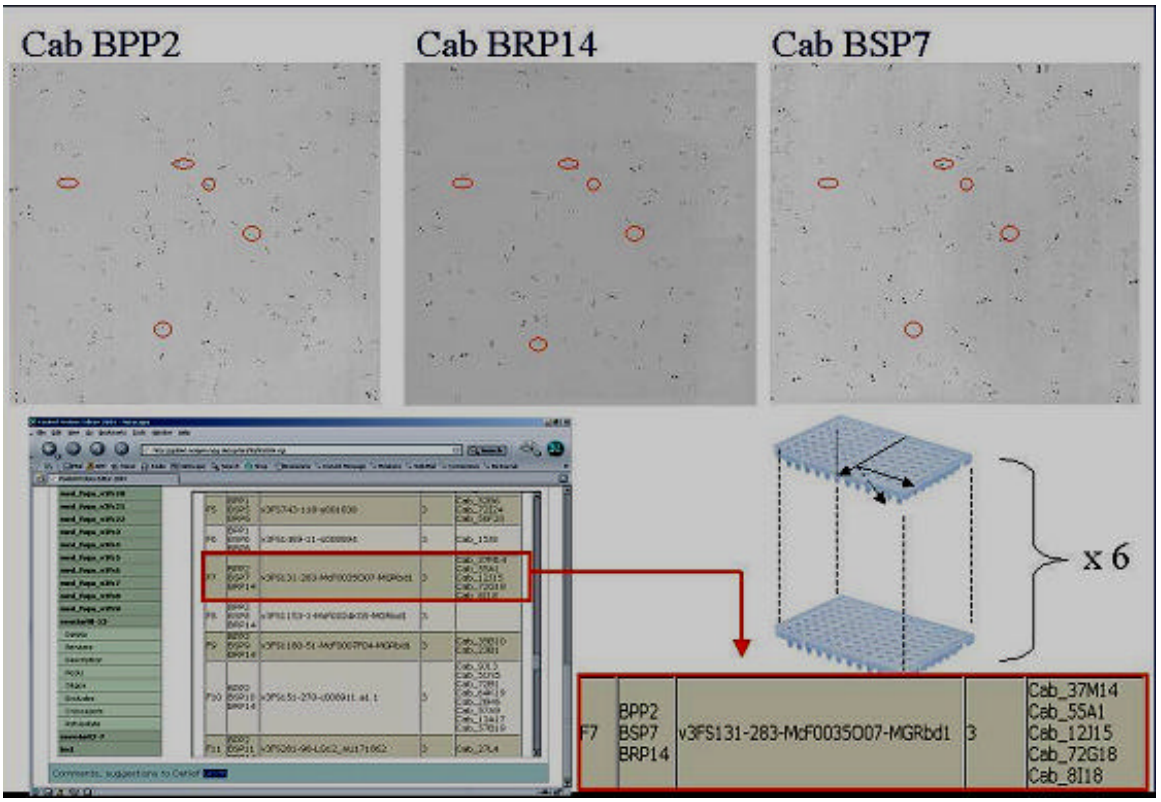

Figure 4-15: Comparison of three BAC Cab filters, which were hybridized with pooled probes BPP2, BRP14 and BSP7 against each other using the PPE software.

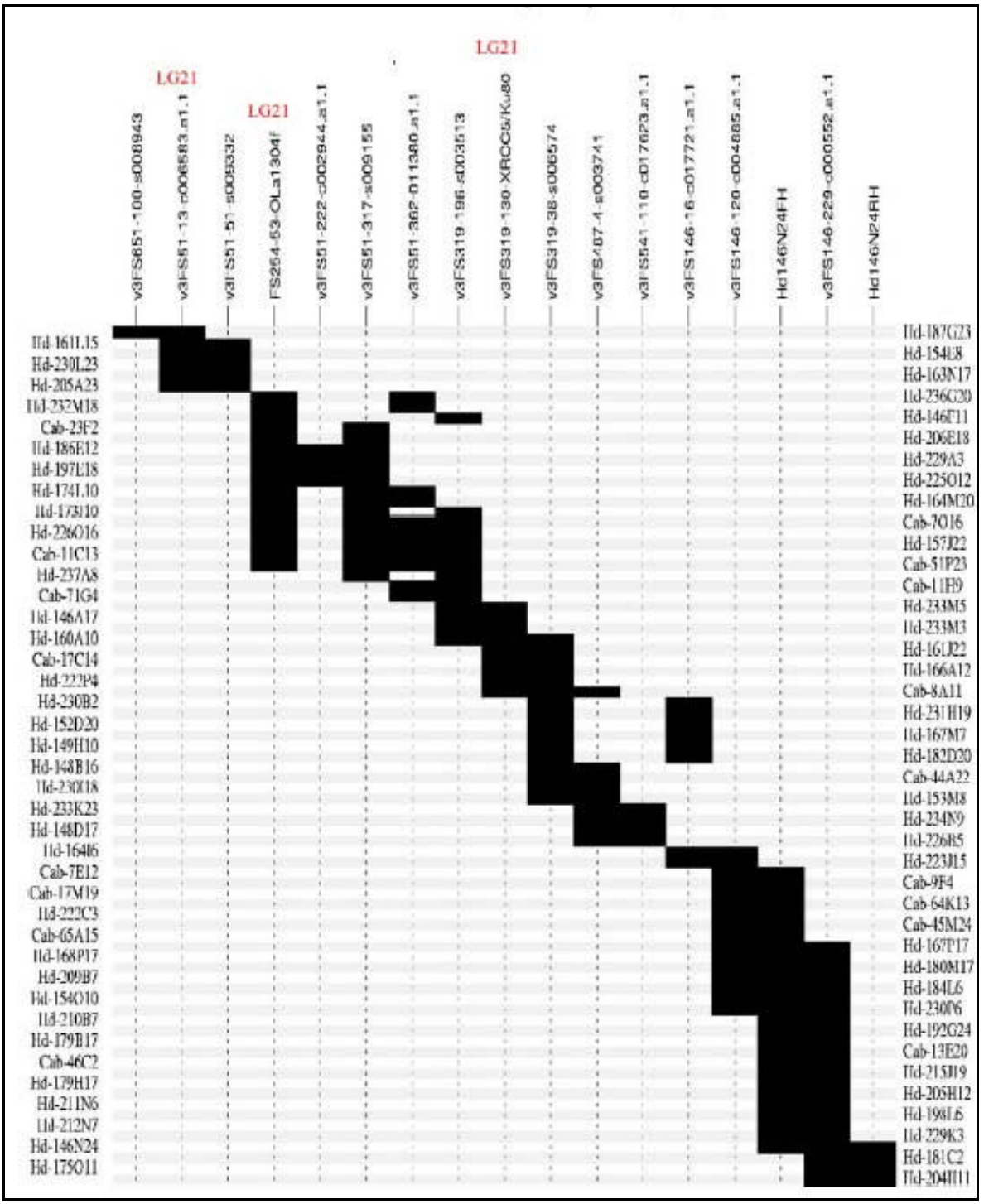

Figure 4-16: Map segment 65 from medaka LG21, containing 17 different markers. Probe v3FS651-100-s008943 (identical to AU169483), OLa1304f and XRCC5/Ku80 are markers that have been genetically mapped to interval 24 of LG21 (see MBase). The segments containing probes v3FS651-100-s008943 and OLa1304f are not connected by BAC clones, but the medaka genetic map and fugu synteny data support co-localization in medaka. The map segment integrates probes that localize to different sequence scaffolds in fugu (scaffolds 651, 51, 319, 487, 541, and 146), and contains one pair of end-probes generated from Hd-rR BAC clone Hd-146N22.

4.5.4 Map construction and physical map

For map construction, we utilized the wprobeorder software (Mott *et al.*, 1993 and Grigoriev *et al.*, 1998). Wprobeorder uses a simulated annealing algorithm to define an optimal order of markers, based on clones that are shared between the markers. The more clones two markers have in common, the closer they are located in the genome. Alternatively, wprobeorder accepts a given list of probes, in the format of a contig file, as input. For the purpose of identifying segments of the medaka genome syntenic to fugu scaffolds, we employed the latter option. Using this approach, a large fraction of the probes immediately fell into contigs, having clones in common with their nearest neighbours (Figure 4-16). Adopting this strategy allowed us to also take advantage of the probes that had hybridized to more clones than expected on average, for map construction. For instance, there are 572 probes in the data set that hybridized to 20 clones or more (Figure 4-11). These results could be due to the presence of closely related gene sequences, e.g. pseudogenes, on other chromosomes that cross-hybridized with the probes used. However, when positive clones were considered only in their local context, multi-match probes could be successfully used, avoiding false links.

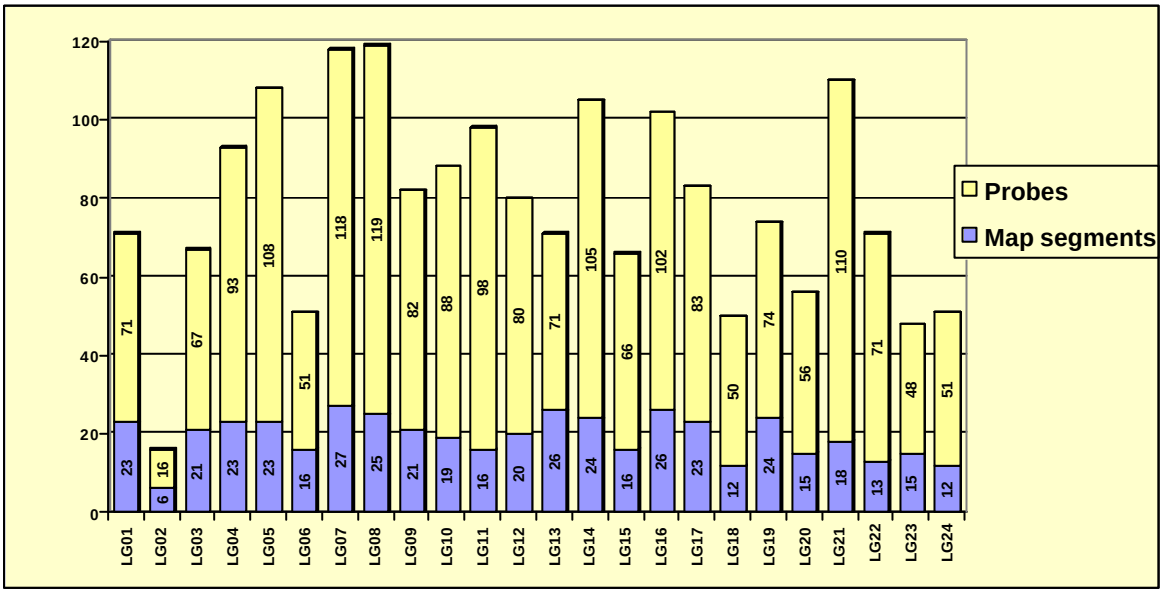

Figure 4-17: Number of map segment located on medaka chromosomes LG1 -LG24 and number of probes contained therein. The majority of our probes with positions on the medaka genetic map are taken from MBase. LG02 map segments are underrepresented, because less genetically mapped markers exist for this chromosome.

In a second step we then refined the probe order manually, if our medaka data suggested a probe path deviating from the marker order in fugu. In regions with low probe densities, gaps in contigs were observed. We kept these contigs together and referred to them as map segments, if genetically mapped markers located within adjacent contigs suggested co-localization (Table 4-2 and Figure 4-17). Once map segments had been established, we then proceeded to inspect links between different map segments and connected them, if supported by genetic map data. This process of map refinement was repeated several times, until all possible links had been evaluated. Finally, we used the increment program (Schalkwyk *et al.*, 2001) to integrate a small set of BAC end-fragment hybridization data into the map. Our final, edited, first generation medaka BAC map consists of 902 map segments, containing a total of 2534 markers from expressed sequences, 133 end-fragment markers from Hd-rR BAC clones and 54 end-fragments from Cab BACs. 462 of map segments are uniquely anchored to a single medaka chromosome (Figure 4-17 and Figure 4-18).

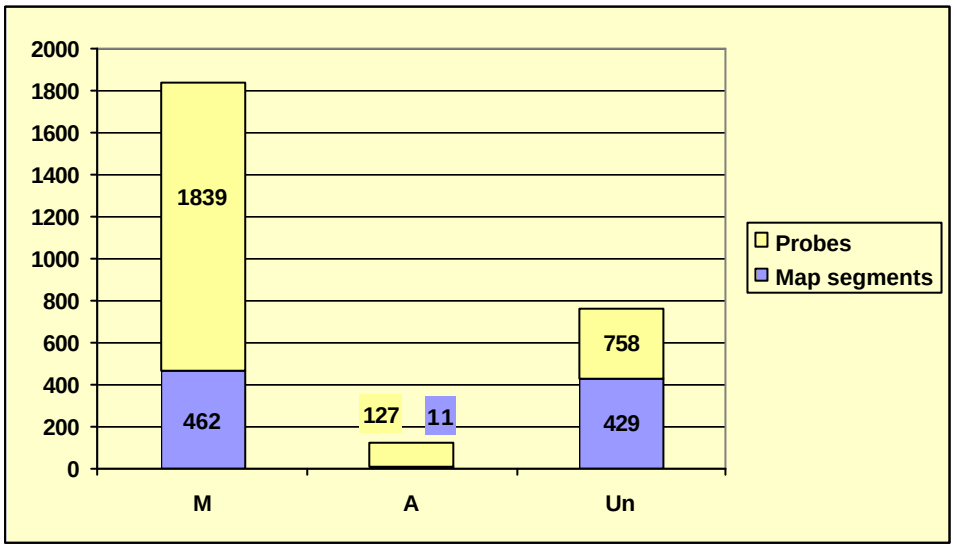

Figure 4-18: Number of map segments that are mapped (M), unmapped (UN), or which are uncertain with respect to their location (ambiguous, A), and total number of probes associated with them.

In a small number of cases, a map segment was anchored to a particular medaka chromosome by multiple genetically mapped markers, but also contained a marker derived from a different linkage group (Figure 4-18). The current data set contains 11 map segments with markers derived from two different medaka linkage groups and one map segment that has markers from three chromosomes integrated. An example is map segment 16, encompassing seven different markers that map to LG21. In addition, probe v3FS10-603-s000478, corresponding to MBase

marker OLa2812c on LG22, is contained within map segment 16. The integration of probe v3FS10-603-s000478 into this map segment is very well supported by several BAC links to adjacent markers. We cannot fully resolve such data conflicts at present. A possible explanation could be the presence of closely related gene copies on different medaka chromosomes, with one copy showing a polymorphism between the mapping strains, resulting in the genetic map assignment. We have also observed fugu scaffolds that cannot readily be translated into medaka BAC contigs. An example is map segment 1, arranged from fugu scaffold 1. In the present map display, six markers localized within 440 kb of fugu scaffold 1 do not have any BAC clones in common. A possible reason for this result could be very different chromosomal organization of such regions in medaka and fugu, for instance, regional expansion in medaka. To clarify such non-trivial situations, we have selected further probes and will incorporate results into more advanced versions of the map. Of the 902 map segments, 439 (49%) are equivalent to singleton contigs, currently not connected to the other map segments. On the other hand, our data set also contains five map segments consisting of more than 20 markers and 35 segments encompassing between 10 and 19 markers. These 40 largest segments alone will cumulatively encompass at least 10% of the medaka genome. We estimate that the average map segment that contains two or more markers has an average marker content of 5.4 and an approximate length of 1 Mb, resulting in a cumulative length of 463 Mb. At the same time, 131 Mb will be contained within the 439 singleton contigs. We, therefore, estimate that the total map length is 594 Mb and approximately corresponds to 74% of the medaka genome. Table 4-2 lists the type and the number of probes used for medaka map construction.

EST-derived, using 35-mer probes	2284
EST-derived, cDNA clone inserts	250
Hd-rR BAC clone end fragments	133
Cab BAC clone end fragments	54
Sum	2721
Sum excluding redundancies	2510

Table 4-2: Marker content of the BAC map.