1 Introduction

This dissertation consists of two parts, 'DNA composition and organization of centromeres' and 'The upper chromosome size limit', both having barley as the common subject.

Barley (*Hordeum vulgare* L.) is an annual cereal of the family Gramineae (grass family), classified in the division Magnoliophyta , class Liliopsida, order Cyperales, family Gramineae. Indications from archaeological remains in the Near East, corresponding geographically to a region extending from Israel through Syria, southern Turkey into Iraq and Iran, suggest that the crop was domesticated about 10,000 years ago from its wild relative *Hordeum spontaneum* (Salamini *et al.* 2002).

Barley is nowadays used commercially for animal feeding, to produce malt for beer and whisky production and for human food applications. It is the fourth most important cereal crop in the world after wheat, rice and maize.

The annual world production of barley (1996-2001) is about 142 million tonnes (http://apps.fao.org/page/form?collection=Production.Crops.Primary&Domain=Production&servlet=1&language=EN&hostname=apps.fao.org&version=default). Barley has a wide range of cultivation and matures even at high altitudes, since its growing period is short, however, it cannot withstand hot and humid climates.

1.1 The centromere: function and structural organization

The centromere is a highly specialized structure of all eukaryotic chromosomes required for correct transmission of the nuclear genetic information from cell to cell and from generation to generation. On monocentric chromosomes it is microscopically recognizable as the primary constriction. It has a central stage role during nuclear

division and fulfils several essential functions. Centromeres are responsible for sister chromatid cohesion until anaphase, represent the site for kinetochore assembly and for attachment of mitotic and meiotic spindle fibres. They are necessary for segregation of sister chromatids into daughter nuclei during mitosis and meiosis II and of homologous chromosomes during meiosis I, and are involved in cell cycle checkpoint control via 'anaphase promoting complex' (for review see Choo 1997; Maney *et al.* 1999).

1.1.1 Centromeric DNA

Although the centromere function is highly conserved among eukaryotes, centromeric DNA sequences are considerably variable between species. A functional centromere of *Saccharomyces cerevisiae* (budding yeast) needs only a 125-bp sequence organized into three elements: CDE I (8 nucleotides), CDE II (an AT-rich ~80-nucleotide sequence) and CDE III (a conserved sequence of 26 nucleotides) (Clarke and Carbon 1985; Hieter *et al.* 1985; Clarke 1990). In *Schizosaccharomyces pombe* (fission yeasts), the central core (*cen1, cen2, cen3*) and at least one block of repeated elements (K-type repeats), has been shown to be essential for correct centromere function (Takahashi *et al.* 1992; Baum *et al*. 1994).

The centromere of higher eukaryotes is usually embedded within large blocks of heterochromatin (White 1973; Choo 1997) characterized by the presence of tandemly repeated DNA in long arrays.

1.1.1.1 Tandem repeats

Many satellite or other tandem repeats with characteristic chromosomal location have been identified and cloned from different organisms. Centromere-associated repeats may represent a considerable fraction of the genomic DNA. Repetitive AT-rich

DNA seems to be a common feature of centromeric DNAs in several organisms such as *S. cerevisiae* (AT-rich CDE II element, Clarke and Carbon 1985; Clarke *et al.* 1993), *Drosophila* (AATAT satellite, Murphy and Karpen 1995; Sun *et al.* 1997), human and other mammals (alphoid DNA with an AT-rich ~171 bp tandem repeat, Manuelidis 1978a, 1978b; Mitchell *et al.* 1985; Willard 1985; Choo *et al.* 1991). Although alphoid satellites are conserved among primates, a considerable variability in sequence became evident even between centromeres of individual human chromosome pairs (Willard 1985; Choo *et al.* 1991; Choo 1997). Similar chromosome-specific variants have been identified in the centromeric minor satellite of the mouse (Kipling *et al.* 1991; 1994).

Various centromere-specific repeats were isolated also from different plant species. For instance *Arabidopsis* centromeres contain tandem arrays of the 180 bp repeat (Martinez-Zapater *et al.* 1986; Simoens *et al.* 1988; Maluszynska and Heslop-Harrison 1991). Species-specific satellite sequences organized in tandem repeats were found also in cereals, e.g. RCS2 in rice (Dong *et al.* 1998), CentC in maize (Ananiev *et al.* 1998), TrsD in rice (Kumekawa *et al.* 2001), the TaiI family in wheat (Kishii *et al.* 2001), CentO in rice (Cheng *et al.* 2002), the pBoKB1 and pBcKB4 repeats in *Brassica* (Harrison and Heslop-Harrison 1995) and the satellite repeat pBV1 in *Beta vulgaris* (Schmidt and Metzlaff 1991). Nevertheless, for some plants (such as field bean and *Tradescantia*) no centromere-specific tandem repeats could be detected (Houben *et al.* 1996).

1.1.1.2 Other centromeric repeats

In addition to the tandemly repeated DNA, a number of other repeat sequences have been found at or near centromeres, which are either genome-wide dispersed or

mainly restricted to centromeric regions, often representing complete or truncated mobile genetic elements, which can be divided into two major groups:

class I including retroviruses (found only in animals); long terminal repeat (LTR) - containing retroelements of the *Ty1*/*copia* and *Ty3*/*gypsy* group, differing in the order of genes encoding their proteins) and non-LTR retrotransposons (e.g. LINE and SINE elements), which transpose by reverse transcription of RNA intermediate, and **class II** (e.g. *Ac*, *En*/*Spm*), which transpose by an excision/insertion mechanism (Kumar and Bennetzen 1999).

In many cases retrotransposons are widely dispersed e.g. *Ty1*-*Ty4* elements inserted into euchromatic regions of *S. cerevisiae* (Boeke 1989; Voytas 1996), *copia* elements present in both eu- and heterochromatic regions in *Drosophila* (Levis *et al.* 1980; Mount and Rubin 1985; Carmena and Gonzales 1995), *Ty1*/*copia* elements in plants (Flavell *et al.* 1992; Brandes *et al.* 1997; Heslop-Harrison *et al.* 1997). Ta elements of *Arabidopsis* (Konieczny *et al.* 1991), the Tnt1 element of tobacco (Grandbastien *et al.* 1997), BARE-1 of barley (Manninen and Schulman 1992; Suonemi *et al.* 1996; 1997) and *Ty/copia* elements in *Vicia* (Pearce *et al.* 1996) and onion (Pich and Schubert 1998) are mainly located in euchromatic regions. Also other elements such as LINEs and SINEs show dispersed chromosomal distribution in human and other mammals (Smit 1996; 2000) and also in plants (Kumar and Bennetzen 1999). Mostly these elements are present in low amount or absent from specific chromosome regions, e.g. centromeres, interstitial and terminal heterochromatin, and rDNA sites (Kumar and Bennetzen 1999). However, there are some exceptions, for example, non-LTR retrotransposon elements *I*, *F*, *G*, and *Doc* are present in the centromeric regions of *Drosophila* chromosomes (O'Hare *et al.* 1991; Pimpinelli *et al.* 1995). The LTR retrotransposon Athila is clustered mainly within pericentromeric heterochromatin (Pélissier *et al.* 1995; 1996) and occupies (together with 180 bp satellite) the centromeric regions of all five

Arabidopsis chromosomes (Pélissier *et al.* 1996; Fransz *et al.* 1998; 2000; Heslop-Harrison *et al.* 1999). Members of the *Ty3*/*gypsy* group of retrotransposons are accumulated within the centromeres of cereals (see below).

1.1.1.3 Centromeric sequences of cereals

Two centromeric sequences were described for cereals. One is the ‚cereal centromeric sequence' (CCS1) family of *Brachypodium* that also occurs in wheat, rye, barley, maize and rice centromeres (Aragón-Alcaide *et al.* 1996) and the other is the *Sau3A9* sequence of sorghum which also hybridized to the primary constrictions of the above species (Jiang *et al.* 1996). Using a barley homologue of *Sau3A9* as a probe, a λ clone (#9) from a genomic library was detected containing a ‚cereba' element (centromeric retroelement of barley). The λ9 clone possesses a complete polygene, with high similarity to the *Ty3*/*gypsy* group of retrotransposons, of which *Sau3A9* represents the integrase encoding region, and flanking sequences similar to CCS1, supposed to represent LTRs of *cereba.* This element hybridized to all barley centromeres (Presting *et al.* 1998). Meanwhile, further conserved sequences representing parts of *gypsy*-like retroelements were found within the centromeres of several cereals such as CentA in maize (Ananiev *et al.* 1998); pHind22 in sorghum, wheat, maize and rye (Miller *et al.* 1998a); RCS1 in rice, rye, barley, sorghum and maize (Dong *et al.* 1998); RCB11 in rice and crwydryn in oats and rye (Nonomura and Kurata 1999; Langdon *et al.* 2000); RIRE7 in rice (Kumekawa *et al.* 1999; Kumekawa *et al.* 2001; Nonomura and Kurata 2001); R11H in wheat (Fukui *et al.* 2001); CRR in rice (Cheng *et al.* 2002). *Gypsy*-like elements (pBv26 and pBp10) were found even within centromeres of dicotyledonous Beta species (Gindullis *et al.* 2001).

The very low conservation of centromeric DNA sequences indicates, that their functional importance is at least controversial, the more so since for several species neocentromeric activities at non-centromeric chromosomal positions have been reported (Depinet *et al.* 1997; du Sart *et al.* 1997). Therefore, it is suggested, that the centromere location might be regulated epigenetically (Vig 1994; Karpen and Allshire 1997).

1.1.2 The kinetochore

The kinetochore is a protein complex associated with eukaryotic centromeres. It plays an important role in interactions of centromeres with the spindle microtubules, in chromosome movements during nuclear divisions, and in the checkpoint (metaphase-anaphase transition) control (Rieder and Salmon 1998; Maney *et al.* 1999). More than 20 proteins associated with the centromere/kinetochore structure have been identified in non-plant organisms. They can be classified into two groups: i) structural proteins (e.g. CENP-A, CENP-B, CENP-C and CENP-H, Sugata *et al.* 2000; Fukagawa *et al.* 2001; for review see Choo 1997), which are constitutively present at centromeres, and ii) passenger proteins (e.g. the BUB family, the MAD family, ZW10, CENP-E, CENP-F and others; Earnshaw and Bernart 1990; Rattner *et al.* 1993; Liao *et al.* 1995; Taylor and McKeon 1997; Starr *et al.* 1997; Yen *et al.* 1991; Yao *et al.* 2000; Saffery *et al.* 2000), which transiently occur at centromeres during nuclear division. Several centromere proteins have been found to be evolutionarily conserved within eukaryotes (Dobie *et al.* 1999). For instance, at least partial homology was found between the yeast Mif2 and the mammalian CENP-C (Earnshaw and Rothfield 1985; Brown 1995) and a putative homologue of maize (Dawe *et al.* 1999). *Drosophila melanogaster* ZW10 homologues are present in *C. elegans*, *A. thaliana*, mice and human (Starr *et al.* 1997). Putative homologs of yeast SKP1 kinetochore protein were found in *Vicia faba* and

barley, and of yeast CBF5p in barley (ten Hoopen *et al.* 2000). Furthermore, cross reactivity was observed for human anti-CENP-E (Yen *et al.* 1991) and anti-CENP-F antibodies (Rattner *et al.* 1993) with kinetochores of *Vicia faba* and barley (ten Hoopen *et al.* 2000).

The high conservation of kinetochore protein sequences, in contrast to the low conservation of centromeric DNA, between remotely related eukaryotic groups suggests that also their functions might be evolutionary conserved.

1.1.3 Aims of the work on barley centromeric DNA

At the beginning of this work large scale sequences and organization of centromeric DNA was not known for plant subjects. Presting *et al.* (1998) have shown, that the sequence of barley λ9 clone possesses in addition to an apparently complete *cereba* element also BARE retroelement sequences, which are dispersed along the chromosome arms of barley (Manninen and Schulman 1992). Furthermore, the *Dra* I restriction pattern of λ9 differed from that of genomic DNA when probed with the barley homologue of *Sau3A9*. For these reasons, it was assumed that this clone might contain either sequences of a centromere-border or represents a chimeric insert not really representative for barley centromeres. Therefore, a genomic BAC library has been screened with the barley homologue of *Sau3A9.* A BAC clone (03J24, here BAC 7) was found to yield FISH signals exclusively at all barley centromeres, and a hybridization pattern comparable to that of genomic DNA after digestion with *Dra* I and Southern hybridization with pGP7 (a barley homologue of the centromere-specific *Sau3A9* element of sorghum) and BCS2 (barley variant of the *cereal centromere sequence*1). This BAC clone was used for further investigations. After shotgun sequencing the aims of this part of the work were:

1. to sequence BAC 7 fully and to align the sequence components for establishing sequence organization characteristic for barley centromeres;
2. to prove whether the CCS1-like sequence belongs to the retroelement *cereba*, and to find out whether other centromere-specific sequences are associated with *cereba*;
3. to compare these sequences with that of other cereal centromeres.

1.2 Chromosome size limitations

The size of chromosomes may vary considerably (from <1 to >20 μm) within and between natural karyotypes. However for theoretical reasons both, lower and upper size limitations must be considered. The question is how such limits are defined.

1.2.1 Lower limit of chromosome size

Indications for a lower size limit for stable chromosome transmission especially during meiosis come from observations made on minichromosomes of yeast, mammals, insects and plants. It was suggested, that in most cases chromosomes should contain ≥ 1% of the host´s genome size for mitotic and clearly more for perfect meiotic stability (for review see Schubert 2001). Possibly, a certain amount of chromatin flanking a centromere is required e.g. for H3 phosphorylation (Houben *et al.* 1999; Manzanero *et al.* 2000) as a lateral support for correct segregation (Schubert 2001). It was recently shown, that in fission yeast flanking heterochromatin is required for cohesion between sister centromeres (Bernard *et al.* 2001). These observations have critical implications for the construction of stable artificial chromosomes.

1.2.2 Upper limit of chromosome size

During nuclear divisions, chromosomes have to be distributed equally to the daughter cells. During anaphase of mitosis and meiosis II chromatids and during meiosis I chromosomes are pulled by spindle fibres attached to centromeres toward the opposite poles of the spindle axis. Later in anaphase a further spindle elongation takes place increasing the distance between poles (Armstrong and Snyder 1989; Hoyt and Geiser 1996). The extension of the spindle is presumably genetically determined (Ming and Hong 2001) although it may vary between specific tissues. Therefore, the extent of spindle axis might be a parameter to determine the upper size limitation for chromosomes.

In *Nicotiana*, abnormally (up to 15-fold) elongated 'megachromosomes' which occured in a few cells of interspecific hybrids (Gerstel and Burns 1966; 1976) could not pass as intact chromosomes from cell to cell, but were broken by the cell plate and yielded chromosome breakage, fragments, dicentrics, rings, anaphase bridges and chromatin elimination. Only the ability to form such megachromosomes was transmitted.

Later on, it was found for *Vicia faba*, that the length of longest chromosome arm must not exceed half of the average length of the spindle axis at telophase (Schubert and Oud 1997). Chromosomes with arms recombinantly elongated beyond this border led to incomplete separation of sister chromatids. As a consequence, breakage of non-separated sister chromatid arms, mediated by the newly forming cell wall during mitosis, caused micronuclei representing chromatin deletions. Viability and fertility of individuals decreased proportionally with the increase of chromosome arm length above half of the average spindle axis dimension, presumably due to a significant increase in apoptotic cells compared to wild-type meristems which is caused by chromatin

deletions and decreases the amount of cells available for tissue differentiation (Schubert *et al*. 1998a).

In *Drosophila*, an abnormally long chromosome C(2)EN with a nearly doubled length of both arms due to both homologs of chromosome 2 sharing a single centromere, caused a ten-fold increase in errors (3.3%) during syncytial embryonic divisions as compared to control embryos (0.3%) (Sullivan *et al*. 1993). This became manifested by chromatin lagging on the metaphase plate, delay of anaphase and final removal of the corresponding nuclei from the population of syncytial nuclei into the inner embryo. Interestingly, in the larval neuroblast cells, the sister chromatids of compound chromosome arms were cleanly separating from each other during late anaphase, most probably because the spindle is longer in the neuroblast cells than in embryonic syncytial nuclei. Although the observed frequency of syncytial mis-division had no obvious impact on viability and fertility of the carrier organism, it seems possible that longer arms might have deleterious effect by further increasing the number of mis-divisions. This indicates that too long chromosome arms may interfere with nuclear divisions also in non-plant organisms.

1.2.3 Aims of the work on upper limit for chromosome arm length in barley

On the basis of previous data, that half of the average length of the spindle axis at telophase defines the upper tolerance limit for chromosome arm length in the field bean, *Vicia faba* (Schubert and Oud 1997), the aims of the second part of this work were:

1. to analyse barley cytotypes with recombinantly elongated chromosome arms as to:
    - mitotic and meiotic spindle axis length
    - separation of sister chromatids into daughter nuclei and formation of micronuclei during mitosis and meiosis

- the impact of elongated chromosome arm(-s) on phenotype and fertility of the plants

2. to test whether the upper tolerance limit for chromosome arm length defined by half of the spindle axis length holds true as a general rule also for other organisms, in this case the monocot barley