WHITE PAPER:

THE INTERNATIONAL BARLEY GENOME SEQUENCING CONSORTIUM (IBSC):

A COORDINATED STRATEGY FOR SEQUENCE ANALYSIS OF THE BARLEY GENOME (*Hordeum vulgare*)

Executive Summary

With an annual harvest of about 140 M tons from ~60 M ha, barley ranks fifth in world crop production. It is a cool season crop grown on all continents. Due to its wide adaptability it is cultivated at higher latitudes, higher altitudes and deeper into semi-arid and arid regions than any other major crop species – thus it is a central component in the majority of agro-ecosystems in temperate regions. Several closely related species within the genus *Hordeum* are adapted to highly diverse and extreme ecosystems around the globe.

The seven chromosomes of barley represent the basic genome of species belonging to the Triticeae tribe. They provide a reference to the genomes of other Triticeae crop species like wheat, rye and some forage grasses. Extensive portions of its genome exhibit strong colinearity to the model grass genomes of rice and Brachypodium.

Building on the genomics resources that have been developed recently, the sequence of the barley genome will give systematic access to virtually any gene underlying a genetically defined trait of interest. The genome sequence will provide the base for advancing barley crop improvement in terms of feed and food quality, pro-biotic utility, grain and biomass yield, resilience to biotic and abiotic stresses, non-food applications as well as nutrient use efficiency. Based on the agricultural importance of barley, improvements in any of these traits will have significant positive effects both economically and environmentally and thus will help meeting future challenges caused by human population growth, climate change and the quest for renewable energy. The availability of an entire genome sequence will enable a leap in speed and efficiency of map-based gene isolation, which still forms the main gateway to the understanding of agricultural traits.

Sequencing the barley genome is an ambitious project. Each chromosome is on average equivalent to two complete rice genomes. Notwithstanding the complexity of the barley genome, the task is manageable and would be achievable over a period of 9 years using state-of-the-art technology and an incremental strategy split into three major phases. The first phase will focus on the development of a physical map and identification of the complete gene repertoire of barley. The second phase will focus on targeted sequencing of prioritized regions of the barley genome while the final phase will be devoted to gap-filling and completion of the genome.

Why sequence the barley genome?

To meet the needs of a growing world population and the challenges imposed by climate change, and to support the development of a future far-sighted bio-based economy, continued efforts are required to develop more effective strategies for agricultural innovation including alternative uses and novel products from existing crop plants. In addition, a sustainable agriculture requires breeding for genetically tailored cultivars to reduce the input of agrochemicals and to adapt crop plants to a changing environment. Generation of a physical map of the barley genome will pave the way towards a deep understanding of small grain temperate cereal genomes and provide a platform for detailed comparative genetic analysis with both established and emerging model plants. It will provide a fundamental resource for isolating genes based on genomic positional information and as such enhance the research communities' ability to investigate and understand the molecular processes underlying phenotypes such as biomass, quality, yield and disease resistance.

Barley ranks fifth in world crop production (harvested area) covering 60 M Ha in 2005 (FAOSTAT 2005). Thus, any investment in this crop species will result in a major impact on worldwide agriculture and economics. In addition, barley combines the least complex (diploid) Triticeae crop genome with a wealth of genetic and genomics resources developed during the past decade. Its genome is highly colinear to other Triticeae species including wheat and rye and also shares numerous agronomic traits with them. Thus, barley is centrally positioned not only as a crop plant but also as a model species for genetic studies within the Triticeae.

Variation among individuals is a fundamental requirement for genetics and has been the driving force behind both quantum change (e.g. improved yield via dwarf varieties) and incremental (e.g. quality) improvements made since crops were first domesticated. Over the past 100 years, conventional barley breeding has resulted in tremendously improved yield and disease resistance. However, further progress in breeding will critically depend on knowledge of genes and on the systematic exploitation of their allelic diversity. The combination of genetic and genomic approaches will lead to a thorough understanding of the basis of major agronomic traits and the underlying metabolic and developmental processes.

In barley, a multitude of genomic regions conferring specific qualities of the crop (such as yield, quality, stress tolerance etc.) has been identified. The underlying genes (and their alleles) need to be found, isolated and studied. Their characterisation will ultimately lead to an understanding of how they exert their effects and to strategies for their efficient utilisation to further improve crop performance. In the majority of cases the most appropriate route to identifying and isolating trait-determining genes will be by their position in the genome. This is time consuming and cumbersome in barley compared to small, fully sequenced genome model plants like Arabidopsis and rice. The majority of traits important for the improvement of cereals having large genomes cannot be studied in these model plants. The availability of a coherent genomic sequence will fundamentally change the current situation, making positional gene isolation in barley both achievable and rapid. We anticipate that many barley genes that are targets for isolation control traits that share their agronomic relevance with wheat and other Triticeae. In many cases, barley will therefore facilitate access to candidate genes on a much wider level. While the annotation of crop plant genomes is still far from being accurate the availability of sequence information from large sections of the barley genome will substantially impact annotation efforts of other plant genomes and shed further light onto functional evolution of grass genomes.

A consortium effort is needed

The complete sequence of the barley genome - at least of its gene space – will be ultimately required to reach a holistic understanding of crop plant performance on the cellular, whole-plant and ecosystem levels.

The scale of a project aiming at sequencing the barley genome (a single chromosome of barley contains one to two genome equivalents of rice) requires strong international collaboration, which is facilitated by the International Barley Sequencing Consortium (IBSC, http://barleygenome.org/). The IBSC has been founded in Adelaide on Sept 1st, 2006, by leading representatives of the following institutions (in alphabetical order):

- Australian Centre of Plant Functional Genomics (ACPFG), AUS
- Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), DEU
- Okayama University (OU), JPN
- National Institute of Agrobiological Sciences (NIAS), JPN
- Scottish Crop Research Institute (SCRI), GBR
- University of California - Riverside (UCR), USA
- University of Helsinki (UH) & MTT Agrifood Research (MTT), FIN
- USDA-Agricultural Research Service (USDA-ARS) at Iowa State University (ISU), USA

The IBSC represents a platform to facilitate the implementation of the agreed research agenda presented below. The consortium aims to coordinate research activities between international research teams, provide assistance in fund raising, and form a strong link to other sequencing initiatives. The policy of IBSC is to achieve timely dissemination of results to a broad public audience.

The following members of the founding institutions were nominated for membership of the steering committee of IBSC (in alphabetical order):

Close, Timothy J (UCR)
Graner, Andreas (IPK)
Langridge, Peter (ACPFG)
Matsumoto, Takashi (NIAS)
Sato, Kazuhiro (OU)
Schulman, Alan (MTT/BI)
Waugh, Robbie (SCRI)
Wise, Roger (USDA-ARS/ISU)

The steering committee is chaired by one of the above listed members of the founding institutions. Presidency will rotate between the committee members on an annual basis.

The founding institutions agreed in a recently signed ‘Memorandum of Understanding’ (MoU) that participating institutions will actively collaborate to achieve the goal of whole genome sequence analysis of barley and mutually support the efforts of IBSC members to acquire additional funding on the national and the international level.

How to approach sequencing the barley genome?

The barley genome will be sequenced by an incremental approach that relies on strategies that have been proven successful with other large genomes (i.e. human[1], maize[2,3]). In brief: a BAC-based high information content fingerprint (HICF) physical map[4] will form the foundation. Initially, this physical map will be anchored to the emerging gene map of barley. Genomic Survey Sequencing (GSS) will provide a resource for further large-scale marker development and enhance anchoring the physical and genetic maps. Anchoring the physical and genetic maps will be further assisted by availability of the rice[5] and *Brachypodium* genome sequence[6], species having genomes largely colinear to Triticeae species. BAC end sequences (BES), accumulated during GSS, will serve as sequence tagged connectors[7] (STC) for supporting and facilitating local contig extension. Initially, whole-genome, random BAC fingerprinting will result in several thousand unconnected contigs. Starting from these, the consortium will focus its efforts to advance the physical maps (reduced number of gaps) of all chromosomes. This will be driven by a combination of community cooperation and individual partner priorities. Chromosomal regions harbouring important traits and regions selected for the study of cross-Triticeae genome dynamics and grass genome colinearity will be emphasized.

I. Developing a physical map of the barley genome

A deep-coverage barley BAC library (15-fold haploid genome coverage, >130 kb average insert size, four enzymes used for cloning one quarter of the library each) will be the common resource for whole genome fingerprinting. The cultivar Morex has been selected as a genomic reference because numerous resources have already been generated using this genetic background. In addition, this will allow the consortium to take advantage of HICF information derived from 60,000 gene-containing BAC clones plus additional 25,000 non-fingerprinted gene-containing BACs from an older community Morex BAC library (NSF-funded project; PI: TJ Close, UCR; M Luo, UC Davis).

A non-redundant set of the previously determined gene-rich clones and the entire new BAC libraries will be HICF fingerprinted and assembled into contigs using available

[1] Human Genome Sequencing Consortium, 2004
[2] Chandler and Brendel, 2002
[3] http://www.nsf.gov/pubs/2004/nsf04614/nsf04614.htm
[4] Luo, et al., 2003
[5] International Rice Genome Sequencing Project, 2005
[6] http://www.jgi.doe.gov/sequencing/why/CSP2007/brachypodium.html
[7] Venter, et al., 1996

software (FPC[8]) to develop an integrated minimum tiling path (MTP). Knowledge of gene content from hybridisation and BES data will provide initial genetic and 'synteny-based' anchors between the physical and genetic maps. It is anticipated that this new effort will generate roughly 1000 -1500 individual contigs per chromosome (7000 – 14,000 contigs per genome) anchored to both the genetic and comparative maps.

II. Anchoring the physical map

A) A high density gene map of the barley genome

The BAC-contig map will be anchored to a high-density gene-based genetic map of barley. Currently, about 4000 barley genes have been mapped to the seven barley chromosomes[9]. This number is expected to increase to ~6000 genes by the end of year 2006. The consortium aims to increase the number of mapped barley genes to set of 15,000 genes, which will become feasible by exploiting the potential of detecting single feature polymorphisms (SFP) and single nucleotide polymorphisms by high-throughput genotyping technologies that are now available[10]. Thus, genetic anchor points for an estimated one third of the gene repertoire of barley will be provided. Markers derived from BAC clones (gene- or repeat-based, see below) will be incorporated into the barley consensus map.

B) Assessing gene content information of BACs

An estimated 1500 mapped barley genes can be already assigned to barley BAC addresses, providing the starting points for anchoring the barley physical to the genetic map. A further set of at least 4000 mapped barley genes (non-cosegregating genes) will be assigned to BAC addresses by one or more deconvolution methods. For example, PCR screening of four-dimensional DNA pools can cover 35 % of the overall library (5.25-fold genome coverage). For this, each gene would require ~400 PCR reactions, for a total of 1.7 million PCR reactions to determining BAC addresses for the whole set of 4000 genes. Other deconvolution methods that are under investigation may simplify this task or increase the number of solved relationships. Alternative methods include the use of combinatorial pools of probes on arrayed BAC MTP clones or combinatorial pools of BACs probed using the Affymetrix Barley1 GeneChip or Illumina GoldenGate assay. A realistic goal is to obtain at least 5000 anchor points between the barley physical and genetic map. Another strategy would rely on large scale skim sequencing of gene-containing contigs (see below) which will provide gene content and thus anchoring information if the above defined target of 15,000 mapped genes can be fulfilled.

[8] Soderlund, et al., 1997; Soderlund, et al., 2000
[9] Stein, et al., 2006; K. Sato, TJ Close, R Waugh, GJ Muehlbauer and A Graner, unpublished data
[10] Rostoks, et al., 2005; Cui, et al., 2005

C) BAC end Sequencing (BES)

Bi-directional BES will be employed in two-thirds of the BAC library resource (10-fold genome coverage) leading to 700000 individual reads accounting for 500 – 600 Mb of the barley genome (8 – 10 % of the barley genome). This will provide first a random snapshot of barley genome organisation and gene density. It is anticipated that ~5% of the sequences will be related to genic sequences and thus allow direct anchoring via homology to: (i) mapped barley genes, (ii) genes discovered on skim-sequenced barley BACs (see below) and (iii) orthologous genes in the rice and *Brachypodium* genome sequences (see below). Due to their direct association with defined BAC addresses, BES derived markers are highly valuable tools for STC development and local contig extension. Previous analyses suggest that up to 10 % of the BAC ends will be suitable for developing repetitive DNA based markers (bridging junctions of nested retrotransposon[11]) also suitable for genetically anchoring of the physical map.

D) Sequence tagged connector strategy for local contig extension

A powerful approach for establishing physical maps and sequencing complex genomes is to use sequence tagged connectors (STC)[12]. Seed BACs originating from loci of interest (i.e. genes of agronomic importance) are skim-sequenced (1 to 2-fold coverage) and compared to a comprehensive source of BES. New, massively parallel sequencing technologies may be ideally suited for this. Homologous BAC ends identify overlapping clones which then enter a contig building routine. Terminal BACs of newly developed contigs are then similarly employed. This strategy will be applied to build on top of the genome-wide HICF approach. Two-thousand BACs (6 % of the barley genome) will be selected and skim-sequenced. The criteria for selecting BACs will take account of consortium priorities (e.g. important trait loci, gene rich regions, genetically anchored contigs). Based on advances in sequencing technology[13] we will also consider shotgun sequencing of larger contigs[14]. The overall objective is to reduce the number of gaps in the whole genome physical map.

E) Colinearity to Rice and *Brachypodium*

Grass genomes share, over large parts of their chromosomes, a high degree of colinearity (synteny). The available gene map demonstrates that across the barley genome a 50 % probability of colinear gene order with rice can be assumed. An even higher level of conserved synteny was observed to several parts of the *Brachypodium*

[11] Paux, et al., 2006
[12] Venter, et al., 1996
[13] Service, 2006
[14] Wicker, et al., 2006

genome[15]. The genome information of both species[16] is already, or will be shortly, available as map-based (rice) or shotgun (*Brachypodium*) sequence within the time-frame of the presented project. These sequences will be employed as scaffolds assisting in the construction of the barley physical map. All information from mapped barley genes and BAC-derived sequences will be superimposed on the genome models based on the rice and *Brachypodium* sequences. This will increase the overall efficiency in developing the barley fingerprint map and will be essential for regions that exhibit insufficient genetic resolution (i.e. large parts of the centromeric regions of all barley chromosomes).

III. Sequencing the Barley Genome

A) Sample sequencing the barley genome

A.1) Genomic Survey Sequencing (GSS) of gene-enriched genomic libraries

Protocols have been developed to enrich fractions of genomic DNA for genic sequences either based on their hypo-methylated state (i.e. methyl-filtration = MF[17]) or on the slower re-association kinetics of low and single-copy sequences compared to repetitive DNA (i.e. C_ot based cloning and sequencing, CBCS, or C_ot filtration = CF[18]). Data from maize showed partial overlap between libraries that had been developed with one or the other method[19]. The pilot gene-enrichment project in maize conducted with high C_ot (HC) and methylation filtered (MF) libraries was the most efficient and effective gene discovery project ever conducted in a higher eukaryote, at the levels of the percentage of genes discovered, the completeness of the gene sequences, and the cost per unit gene discovery[20]. A pilot study indicated 18-fold enrichment of genic sequences for barley[21]. Deep-coverage gene-enrichment libraries will be randomly sequenced for up to 500000 reads per approach, thus generating up to 500 Mb of barley genomic sequence (10 % of the barley genome) with a >90 % probability to tag each barley gene by at least a single GSS read.

Because gene-enriched libraries provide access to up- and downstream sequences of genes this dataset will be highly complementary to the 500000 barley ESTs present in public databases. GSS information in combination with the barley EST resources and BES data will provide a refined estimate of total gene content in barley.

[15] Griffiths, et al., 2006
[16] International Rice Genome Sequencing Project, 2005; http://www.jgi.doe.gov/sequencing/why/CSP2007/brachypodium.html
[17] Rabinowicz, 2003
[18] Peterson, et al., 2002
[19] Springer, et al., 2004
[20] Springer, et al., 2004; Barbazuk, et al., 2005; Rabinowicz, et al., 2005
[21] Rabinowicz, et al., 2005

A.2) Contig sequencing

The gene-containing portion of barley is estimated to represent less than 10 % of the total genome, distributed among BACs comprising about 40% of the genome. About five hundred 1Mb (average) contigs (10% of the barley genome) will be selected based on partial knowledge of their gene content and genetic location and sequenced to a 5-fold coverage (Phase I sequence quality). Resulting sequence assemblies will contain a varying number of gaps. For most clones, the relative orientation of fragments will be deduced by sequence comparison to rice and Brachypodium and the finishing of BAC sequences will be avoided. A smaller number of large contigs will be advanced to finished sequence based on the priorities of the international partnerships (e.g. important trait loci, comparative grass genomics).

B) Genomic sequencing

B.1) Targeted chromosome sequencing

Providing a test case, a barley chromosome will be targeted for developing a gap-free (as much as achievable) physical map. This will provide the backbone for clone-based sequencing of a whole chromosome including the larger heterochromatic regions neighbouring the centromere. The organisation of a large cereal genome centromere will be revealed and the distribution and frequency of occurrence of genes can be disclosed for these regions of the barley genome. It is unclear from a current standpoint to what extent further sequencing of similar regions of other barley chromosomes will be essential for complete understanding of overall barley genome structure and evolution.

B.2) Whole genome sequencing

The overriding goal of IBSC is a finished whole genome sequence of the barley genome. This task needs to be considered in the context of rapid, dynamic advances in sequencing technology. In view of recently introduced novel sequencing technology that allow generation of 50 – 200 Mb of genomic sequence in a single day[22], either ordered-clone-based whole-genome sequencing or whole-genome shotgun (WGS) sequencing will be attempted within the context of the barley project. A whole-genome sequence will bring together and take advantage of all previously developed information: the physical map, BES and GSS and local contig sequence data.

[22] Service, 2006; Metzker, 2005

IV. Bioinformatics for sequencing the barley genome

A) Genome annotation

A federated bioinformatics infrastructure will be established, to advance the physical map assembly, sequence and genome annotations, to support database curation and generally exploit publicly available genomics data information. This will build on expertise that is already available within the consortium[23]. To this end, tight links to related genome initiatives, especially those aiming at sequencing and annotating grass genomes (rice, maize, Brachypodium, wheat) have been established and will be strengthened. Available tools originating from various international genome physical mapping and sequencing projects will be adopted for barley (e.g. BES annotation pipeline established for maize, MIPS[24]). Furthermore, the specific aspects of Triticeae sequence and genome organisation will require the development of new tools for the annotation of gene structure and the plethora of different repetitive DNA elements, the latter will undoubtedly complicate sequence assembly. Therefore, barley genome annotation will be in close coordination with the TIGR cereal genome annotation group[25] (led by R Buell), the Rice Genome annotation Group of IRGSP, and the *Brachypodium* annotation group led by the US Department of Energy Joint Genome Institute. The accumulated information will be connected wherever possible to public genomics resources available for other grass species (the complete sequence of rice, maize, sorghum, Brachypodium and survey sequences of the wheat genomes) within a data warehouse environment for Triticeae genomics. This will provide essential cereal crop information for application in research and applied breeding.

[23] Bioinformatics units at UCR, ISU, SCRI, IPK, NIAS/IRGSP; close interaction with MIPS, TU Munich/GSF
[24] Messing, et al., 2004
[25] http://www.wheatgenome.org/id33.html

Concomitant activities by partners of IBSC

USA

UC Riverside, UC Davis (PI: TJ Close, M Luo)

An attempt to generate BAC contig maps for the barley gene space utilized BACs identified by hybridisation of an older Morex BAC library with overgo probes for 12500 genes. The results, when added to all prior screening of this library, provided a cumulative total of 84000 gene-positive BAC addresses. Approximately 1500 genes represented by the collective list of probes are already anchored to the genetic map of barley and further deconvolution of gene / BAC address assignments is underway. 60,000 of the above indicated clones have been analysed by HICF (6600 contigs, 4400 singletons; TJC and ML unpublished data). The resulting MTP is available and will be integrated into the whole genome HICF map of barley.

Status: Funded

USA/GBR/DEU

UC Riverside, U Minnesota, SCRI Dundee, IPK (PI: TJ Close, G Muehlbauer, R Waugh / L Ramsay, A Graner)

Oligo-Probe Assays (Illumina Golden Gate Assay) for over 4000 barley genes have been designed for high throughput mapping and genotyping leading to genetic map positions of 3000+ genes in a genetic consensus map. Further genes could be integrated into the same map via genetic association.

Status: Funded

DEU/AUS

IPK Gatersleben/ACPFG Adelaide/TU Munich (PI: A Graner / N Stein, P Langridge, HW Mewes / KFX Mayer)

Developing of a contig map of the barley genome by HICF of 350,000 clones has been initiated. A deep coverage (10-fold haploid genome equivalents) multi-enzyme BAC library is under development at ACPFG for becoming transferred to IPK for HICF on ABI 3730xl technology.

Status: Funded

JAP

U Okayama / NIAS (PI: K Sato / T Matsumoto)

Develoment of 35000 full length cDNAs (fl-cDNA) is in progress. The data will be integrated into the IBSC project and thereby facilitate full gene annotation in barley and related grass species. 3000 barley genes have been mapped to the barley linkage

groups at low resolution. Access by the IBSC project to the latter information is granted. 500 BAC clones (from cv H. nijo) genetically anchored to chromosome 3H have been selected and are subject of complete sequencing. This information will reveal soon a comprehensive dataset for comparative genomics to rice, *Brachypodium* and wheat. It will provide an excellent dataset for studying intra-species genome colinearity by comparison to orthologous sequences derived from cv Morex.
Status: Funded

FIN
MTT Agrifood, U Helsinki (PI: A Schulman)
Precise knowledge of many classes of barley repetitive DNA elements is available leading to development of several repeat-based marker techniques. This know-how will greatly contribute to exploit BES and GSS data for marker development and map anchoring. In addition, an in-house transposable element annotation and a BAC Sanger sequencing pipeline as well as newly installed 454/Roche GS20 (upgrading to GS100) and local Illumina station are available. A proposal was submitted to evaluate 454 sequencing technology for analysis and annotation of repetitive DNA in Triticeae.
Status: Proposal submitted, decision expected by early 2007.

GBR/ITA/DEU
SCRI, U Udine, IPK (PI: R Waugh, M Morgante, N Stein)
A project was submitted to the 1st call of "European Research Area – Plant Genomics (ERA-PG)" aiming at the bidirectional sequencing of 350,000 BACs of the BAC library that is being employed for developing the barley physical map.
Status: Recommended for Funding

USA
TIGR, UC Riverside, U Minnesota, Iowa State, UC Davis, U Georgia (coordinating PIs: T Close/R Wise)
A proposal (TRPGR: Sequencing the barley gene-space) was submitted to NSF-Plant Genome Program (Oct 2006) aiming at deep coverage GSS utilizing gene enriched libraries (MF and CF) and sequencing nine Mb-sized contigs of barley BACs.
Status. Proposal submitted, decision expected by early 2007.

AUS
ACPFG Adelaide, Murdoch University (coordinatin PIs: P Langridge / R Appels)
A proposal has been lodged with the Australian Government to provide AUS$5 million over 3 years to support work towards wheat and barley genome analysis and

sequencing. Approximately 50% would be made available to support the IBGS activities.
Status: Proposal submitted, decision expected by early 2007.

Data management and release policy

The aim of IBSC is to ensure for direct dissemination to the public of all data accumulated in course of physical map development and sequencing of the barley genome. A website (http://barleygenome.org/) will provide up-to-date documentation of the status and progress within the diverse tasks of IBSC. A database for physical map assembly, status of physical/genetic map anchoring and sequence/genome annotation is under construction according to current standards. Sequencing data generated through activities of IBSC will be immediately submitted to NCBI Genbank, EMBL and DDBJ. Physical map and anchoring information (mapped genes / markers) will be provided to alternative genome related databases databases as well (NCBI, GrainGenes, Gramene, TIGR, Harvest:Barley, The Barley Physical Mapping Database, Barley CAP *Hordeum* Toolbox).

Interaction with related activities

Close interactions with related activities have been established to promote an efficient transfer of know-how and exchange of information with other sequencing projects. Under the auspices of the International Triticeae Mapping Initiative (ITMI, http://wheat.pw.usda.gov/ITMI) a vital informal exchange of research in the Triticeae and related grass species has a long lasting tradition at the international level. Recently, a European Triticeae Genomics Initiative (ETGI, http://ww.etgi.org) was established for coordinating Triticeae genomics activities at the EU scale and for developing a tight network of research collaboration. The coordinating functions of ETGI will be supported by TritiGen, a broad COST action (F&A0604), which starts in early 2007. Furthermore, tight interaction in aspects of HICF map construction, comparative genome sequencing and genome annotation is underway with the Wheat D-Genome project (PI: J Dvorak, UC Davis) and the wheat chromosome 3B project (PI: C Feuillet, INRA Clermont).

Meetings, Coordination and Collaborations within IBSC

Progress in the aims of IBSC will be presented annually at the Plant and Animal Genome Conference in San Diego, California. Additional regular meetings will be held at the summer workshop of ITMI which is organised at changing locations.
The IBSC will invite experts in the field of physical map construction, genome sequencing and genome sequence data basing to join a scientific advisory board (SAB, 3 members) for accompanying and consulting the progress made by IBSC and collaborators.
Members of the IBSC Steering committee will meet biannually to achieve closely coordinated activities among members and to funding bodies of the participating countries – the members of the SAB will be invited to participate. IBSC and cooperating partners will raise funds for appointing a manager of IBSC activities (website, PR, workshop organization, coordination of grant applications, negotiations with funding bodies).

The founding member institutions of IBSC foster intensive interactions among each other and with the international Triticeae research community. IBSC gathers together institutions with strong and long lasting track records in barley and other genome analyses, combining all necessary expertise to commence the proposed activities. Resources for barley genome analysis have been intensively shared during the past decade between partners in Europe, USA, Australia and Japan.

Funding scheme

To implement the above mentioned research agenda, substantial additional funding will be required. According to our present estimate, the complete sequence of the genome of barley would amount to about 50 M €. The largest part of this figure is attributed to DNA sequencing and we expect that depending on technology development this figure may be subject to substantial changes. While until recently, despite long lasting international collaboration at the level of individual research groups, genome research in barley was funded mainly on the national level, both IBSC, ETGI and the European Barley Genome Net have generated the platform for the acquisition of EU-funding in the ERA-NET PG and in the FP7 programme. Despite the opportunities opened up by the EU, successful completion of the project will rely on national funding. Therefore, all IBSC members are committed to apply for funding in a complementary non redundant way, to warrant an efficient use of the financial resources and to maximize the output of their efforts.

A) Prospected financial volume[26]

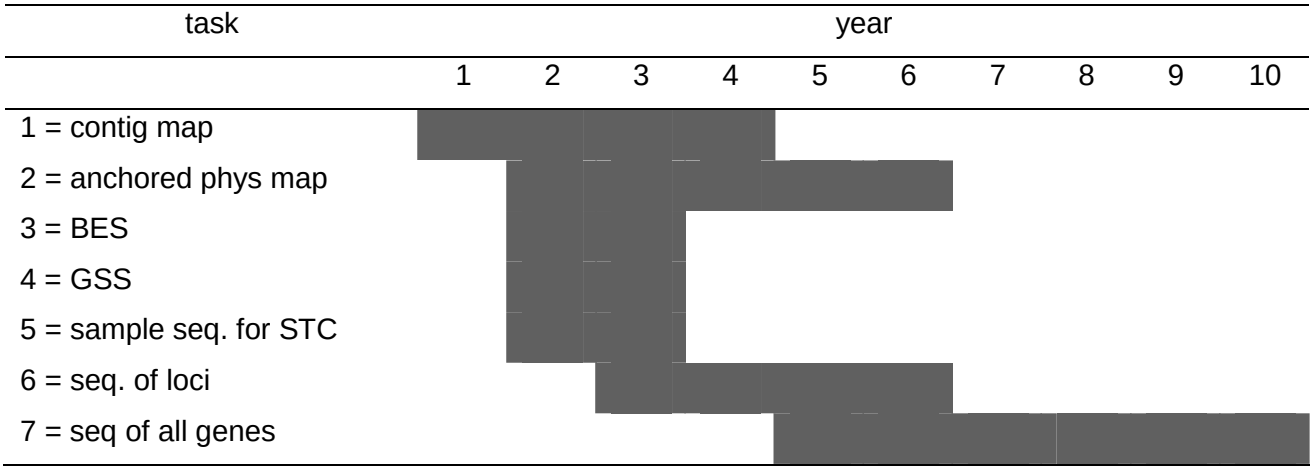
Task 1: Contig map construction (HICF) (15 x coverage)	3.0 M EUR
Task 2: Anchoring of physical map	2.0 M EUR
Task 3: BAC end sequencing (700000 reads)	1.0 M EUR
Task 4: GSS (1 Mio reads MF/CF)	1.5 M EUR
Task 5: Sample sequencing of BACs/ contigs for STC (2000 BACs)	5 – 10 M EUR
Task 6: sequencing gene content of important loci (500 contigs, 1 Mb each)	2 – 5 M EUR
Task 7: Sequence of the barley gene space[27] (5 x coverage of 1 Gb = 20% of the genome)	costs largely depending on technology development

B) Milestones / Timeline

IBSC assumes that developing a sequence ready physical map and complete sequencing of the barley gene space will last over a period of about 10 years starting from year 2006. The final duration required for such an effort will depend on overall funding, the pace of advancement in sequencing technology and on adaptation of the technologies to sequencing complex plant genomes carrying large quantities of repetitive DNA.

[26] costs will largely depend on advances in sequencing technology development
[27] based on estimated 2500 EUR for 5x coverage Sanger sequencing of 100 kb (average BAC)

Gantt Chart of IBSC planned tasks for barley genome sequencing



Citations:

1. Human Genome Sequencing Consortium, *Nature* 431, 931 (2004).
2. V. L. Chandler, V. Brendel, *Plant Physiology* 130, 1594 (2002).
3. M.-C. Luo *et al.*, *Genomics* 82, 378 (2003).
4. International Rice Genome Sequencing Project, *Nature* 436, 793 (2005).
5. J. C. Venter, H. O. Smith, L. Hood, *Nature* 381, 364 (1996).
6. C. Soderlund, I. Longden, R. Mott, *Comput. Appl. Biosci.* 13, 523 (October 1, 1997, 1997).
7. C. Soderlund, S. Humphray, A. Dunham, L. French, *Genome Res.* 10, 1772 (November 1, 2000, 2000).
8. N. Stein *et al.*, *Theoretical and Applied Genetics*, under revision (2006).
9. N. Rostoks *et al.*, *Genome Biology* 6, R54 (2005).
10. X. Cui *et al.*, *Bioinformatics* 21, 3852 (October 15, 2005, 2005).
11. E. Paux *et al.*, *Plant Journal* in press (2006).
12. R. F. Service, *Science* 311, 1544 (March 17, 2006, 2006).
13. T. Wicker *et al.*, *BMC Genomics* 7, 275 (2006).
14. S. Griffiths *et al.*, *Nature* 439, 749 (2006).
15. P. D. Rabinowicz, *Methods Mol Biol* 236, 21 (2003).
16. D. G. Peterson *et al.*, *Genome Res.* 12, 795 (May 1, 2002, 2002).
17. N. M. Springer, X. Xu, W. B. Barbazuk, *Plant Physiol.* 136, 3023 (October 1, 2004, 2004).
18. W. B. Barbazuk, J. A. Bedell, P. D. Rabinowicz, *BioEssays* 27, 839 (2005).
19. P. D. Rabinowicz *et al.*, *Genome Res.* 15, 1431 (October 1, 2005, 2005).
20. M. L. Metzker, *Genome Res.* 15, 1767 (December 1, 2005, 2005).
21. J. Messing *et al.*, *PNAS* 101, 14349 (October 5, 2004, 2004).