A Physical Map of the 1-Gigabase Bread Wheat Chromosome 3B
Etienne Paux, *et al.*
Science **322**, 101 (2008);
DOI: 10.1126/science.1161847

35. E. R. Wood, P. A. Dudchenko, H. Eichenbaum, *Nature* **397**, 613 (1999).
36. C. M. Bird, N. Burgess, *Nat. Rev. Neurosci.* **9**, 182 (2008).
37. We thank the patients for their cooperation and participation in this study. We also thank E. Ho, B. Scott, E. Behnke, R. Kadivar, T. Fields, A. Postolova, K. Laird, C. Wilson, R. Quian-Quiroga, A. Kraskov, F. Mormann, and M. Cerf for assistance with data acquisition; B. Salaz and I. Wainwright for administrative help; and I. Kahn, Y. Nir, G. Buzsáki, E. Pastalkova, and S. Gilaie-Dotan for discussions and comments on this manuscript. This work was supported by NINDS (to I. Fried), Israel Science Foundation (to R. Malach), Binational United States–Israel grant (to I. Fried and R. Malach), and Human Frontier Science Program Organization (HFSPO) fellowship (to R. Mukamel).

A Physical Map of the 1-Gigabase Bread Wheat Chromosome 3B

Etienne Paux,[1] Pierre Sourdille,[1] Jérôme Salse,[1] Cyrille Saintenac,[1] Frédéric Choulet,[1] Philippe Leroy,[1] Abraham Korol,[2] Monika Michalak,[3] Shahryar Kianian,[3] Wolfgang Spielmeyer,[4] Evans Lagudah,[4] Daryl Somers,[5] Andrzej Kilian,[6] Michael Alaux,[7] Sonia Vautrin,[8] Hélène Bergès,[8] Kellye Eversole,[9] Rudi Appels,[10] Jan Safar,[11] Hana Simkova,[11] Jaroslav Dolezel,[11] Michel Bernard,[1] Catherine Feuillet[1]

As the staple food for 35% of the world's population, wheat is one of the most important crop species. To date, sequence-based tools to accelerate wheat improvement are lacking. As part of the international effort to sequence the 17–billion–base-pair hexaploid bread wheat genome (2*n* = 6*x* = 42 chromosomes), we constructed a bacterial artificial chromosome (BAC)–based integrated physical map of the largest chromosome, 3B, that alone is 995 megabases. A chromosome-specific BAC library was used to assemble 82% of the chromosome into 1036 contigs that were anchored with 1443 molecular markers, providing a major resource for genetic and genomic studies. This physical map establishes a template for the remaining wheat chromosomes and demonstrates the feasibility of constructing physical maps in large, complex, polyploid genomes with a chromosome-based approach.

Among plants providing food for humans and animals, one of the oldest and most widespread is wheat (*Triticum aestivum* L.). Despite its socioeconomic importance and the challenges that agriculture is facing today (*1*), wheat genomics and its application to crop improvement are lagging behind those of most other important crops. The wheat genome has always been viewed as impossible to sequence because of its large amount of repetitive sequences (>80%) and its size of 17 Gb, which is five times larger than the human genome. The largest wheat chromosome (3B) alone is more than twice the size of the entire 370-Mb rice genome (*2*), whereas the entire maize genome (2.6 Gb) is about the size of three wheat chromosomes (table S1). Further complicating the challenge, bread wheat is a relatively recent hexaploid (2*n* = 6*x* = 42) containing three homoeologous A, B, and D genomes of related progenitor species, meiotic recombination is not distributed homogeneously along the chromosomes, and intervarietal polymorphism is very low.

Genome sequencing is the foundation for understanding the molecular basis of phenotypic variation, accelerating breeding, and improving the exploitation of genetic diversity to develop new crop varieties with increased yield and improved resistance to biotic and abiotic stresses. These new varieties will be critical for meeting the challenges of the 21st century, such as climatic changes, modifications of diets, human population growth, and the increased demand for biofuels. Physical maps are essential for high-quality sequence assembly regardless of the sequencing strategy used [such as bacterial artificial chromosome (BAC)–by–BAC or whole-genome shotgun strategies], and they will remain pivotal for de novo sequencing even with the advent of short-read technologies (*3*). As the foundation for genome sequencing, physical maps have been established for a dozen plants species so far, including cereals such as maize, rice, and sorghum (*4–6*). Recently, the development of new genomic resources for analyzing wheat paved the way for physically mapping and ultimately sequencing a species for which this was unthinkable a few years ago.

A physical map with 10-fold coverage of the 17-Gb bread wheat genome would require more than 1.4 million BAC clones to be finger-printed, assembled into contigs, and anchored to genetic maps. Although whole-genome BAC libraries are available and fingerprinting millions of BAC clones is technically feasible with high-information-content fingerprinting (HICF) (*7*), assembly to accurately depict individual chromosomes and the anchoring of homoeologous BAC contigs onto genetic maps remains daunting. To address these issues, we used a chromosome-based approach (*8*) to construct a physical map of the largest hexaploid wheat chromosome (3B) and, to compensate for the inherent limits of the wheat genome (the lack of recombination and polymorphism), we deployed a combination of genetic mapping strategies for anchoring the physical map.

Fingerprinting and contig assembly were performed with BAC clones originating from sorted 3B wheat chromosomes of Chinese Spring (*9*), the reference cultivar chosen for genome sequencing by the International Wheat Genome Sequencing Consortium because of its previous use for cytogenetic studies and the availability of a set of aneuploid lines (*10*). 67,968 3B BAC clones were fingerprinted with a modified (*11*) HICF SNaPshot protocol (*7*), and a total of 56,952 high-quality fingerprints (84%) was obtained. A first automated assembly (*11*) resulted in a final build of 1991 contigs with an average size of 482 kb for a total length of 960 Mb (table S2). One hundred ninety-seven contigs were larger than 1 Mb; the largest was 3852 kb in size. A minimal tiling path (MTP) consisting of 7440 overlapping BAC clones was defined for further analyses. After the preliminary automated assembly, contigs were merged manually (*11*), resulting in a final assembly of 1036 contigs with an average size of 783 kb (table S2) covering 811 Mb (~82%) of the estimated 995 Mb (*12*) constituting chromosome 3B.

Contig assembly was validated through BAC library screening with markers derived from BAC-end sequences (BESs) (*13*) and by genetic mapping (*11*). Out of 421 markers derived from BESs, 369 (88%) correctly identified the BAC clones belonging to computationally identified contigs. Conversely, 35 markers originating from the same contigs mapped to the same genetic

[1]Institut National de la Recherche Agronomique, Université Blaise Pascal (INRA-UBP), UMR 1095, Genetics Diversity and Ecophysiology of Cereals, Clermont-Ferrand, France. [2]Institute of Evolution, University of Haifa, Haifa, Israel. [3]Department of Plant Sciences, North Dakota State University, Fargo, ND, USA. [4]Commonwealth Scientific and Industrial Research Organization Plant Industry, Canberra, Australia. [5]Agriculture and Agri-Food Canada, Cereal Research Centre, Winnipeg, Canada. [6]Diversity Arrays Technology, Yarralumla, Canberra, Australia. [7]INRA–Unité de Recherches en Génomique-Info, Versailles, France. [8]INRA–Centre National de Ressources Génomiques Végétales, Toulouse, France. [9]International Wheat Genome Sequencing Consortium, Eversole Associates, Bethesda, MD, USA. [10]Centre for Comparative Genomics, Murdoch University, Perth, Australia. [11]Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Olomouc, Czech Republic.

locus. The wheat genome has a high content of long terminal repeat retrotransposons (> 67%) (13) and a large number of tandemly repeated sequences, which may result in the misassembly of BAC clones. A statistical analysis (11) indicated that less than 10% of the fragments were randomly shared between any two non-overlapping BAC clones. We calculated that approximately 58% overlap between fingerprints was required for automated assembly (fig. S1), and therefore with the high stringency (Sulston score = le = 75) used for the assembly (11), the repeated sequences did not affect the quality of the contig build.

The physical map was anchored with 685 microsatellites [simple sequence repeats (SSRs)], some of which were designed from the BESs, as well as expressed sequence tag (EST) markers (table S3) previously mapped to chromosome 3B (14) and identified on the basis of their synteny with rice chromosome 1. In total, 291 SSR markers were anchored to 203 contigs representing 219 Mb, and 394 ESTs were anchored to 250 individual contigs representing 283 Mb (Fig. 1 and table S3). Four hundred seventy-two additional contigs representing 452 Mb were then anchored with 711 insertion site–based polymorphism (ISBP) markers derived from 19,400 BESs (13) (Fig. 1 and table S3). We also tested the multiplexed and genomewide Diversity Arrays Technology (DArT) (15) by hybridizing a wheat array composed of 5000 DArT markers with three-dimensional pools of the MTP (11). Thirty-five DArTs were unambiguously assigned and anchored to 25 individual contigs (19 Mb) (table S3).

In total, 1443 molecular markers were linked to 680 BAC contigs representing 611 Mb and 75% of the 3B physical map. The longer contigs were anchored more easily (fig. S2), and more than 80% of contigs longer than 900 kb were anchored. Very few contigs (50, <80 Mb) were anchored with all marker types, and most (463 out of 680) were anchored by a single marker type (Fig. 1). This indicates that the different classes of markers cover different regions of the genome and should be used in combination to ensure optimal representation of the chromosome.
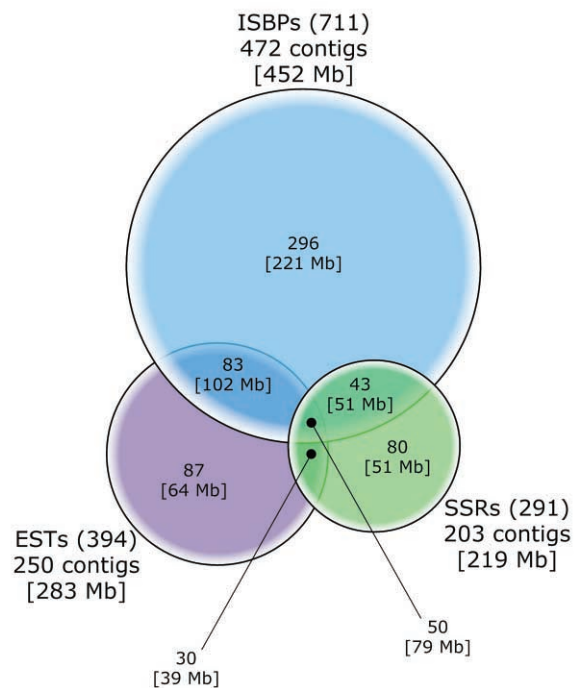
The anchored physical map can expedite map-based cloning, but its full value is achieved by determining the relative contig order along the chromosome and producing an integrated physical map. To integrate the 3B physical map, we used deletion mapping [in which the absence of two genetic sites in the same deletion line (that is, one of many lines containing small deletions in specific sites along the chromosome) is a measure of the maximal distance between them] and meiotic mapping (in which the relative order of markers is determined on the basis of recombination). This combined approach was necessary because in wheat, the resolution of meiotic mapping is limited by both the nonhomogeneous distribution of recombination events along the chromosome arms (for example, on chromosome

3B, 42% of the physical map length is represented by only 2.2% of the genetic map length in the centromeric regions) and the low level of polymorphism in the cultivated pool.

Deletion mapping resulted in the integration of 599 contigs (556 Mb, ~56% of the chromosome) in 16 physical intervals along chromosome 3B (table S3). To assess whether our physical map accurately depicts 3B, we systematically compared the total size of the contigs present in the 16 intervals defined by genetic deletions (so-called deletion bins) with the size of the genetic deletion as estimated cytologically from the chromosomes in the corresponding deletion line stocks (16) (table S3). Coverage within the bins ranged from 33 to 99%, with an average of 56% (fig. S3). The most terminal bin on the short arm (bin 3BS3-0.87-1.00) was only 10% covered and had a smaller average contig size (630 kb) than other bins (957 kb), suggesting that the telomeric and heterochromatic region of the short arm may be underrepresented in our BAC library.
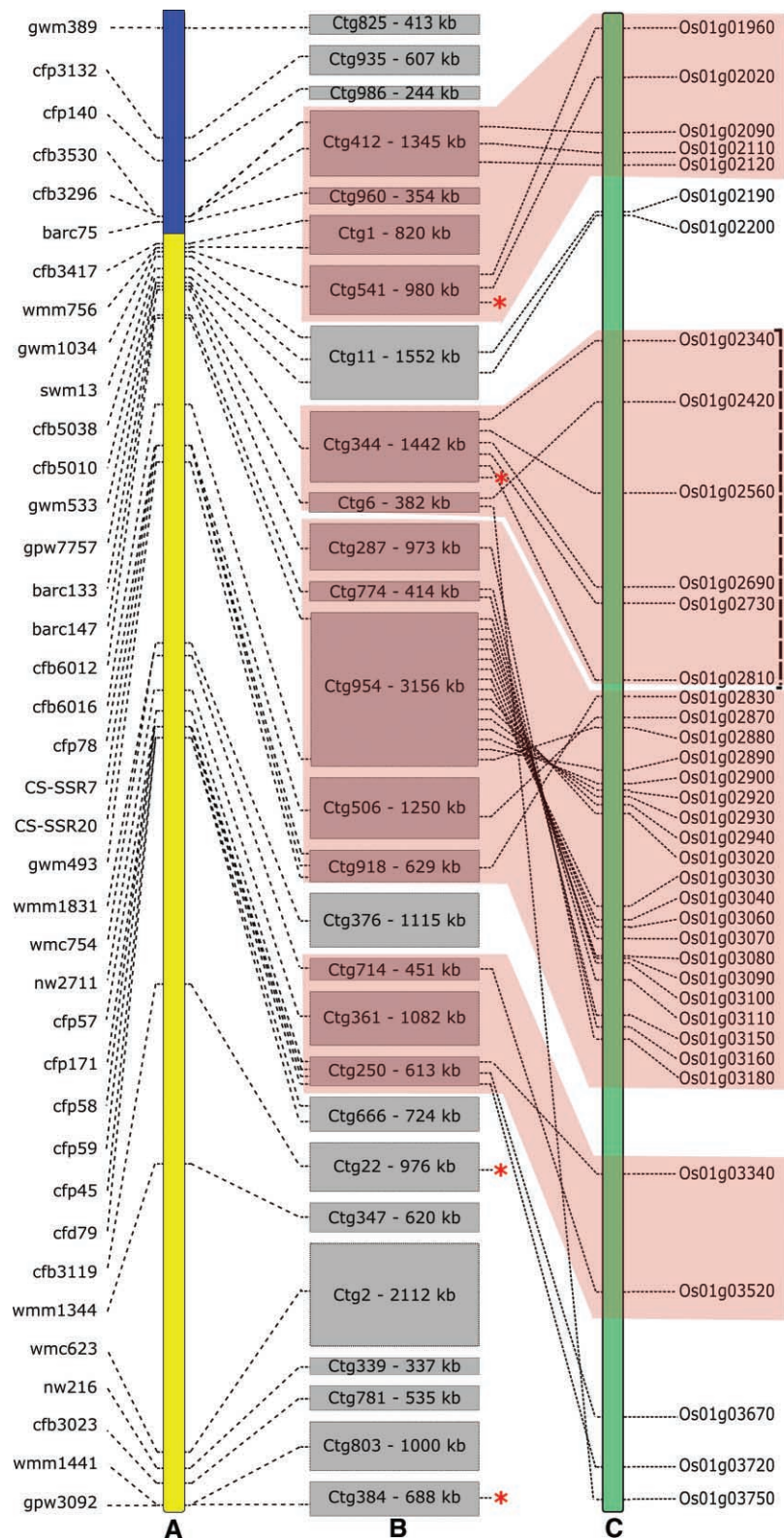
We tested high-resolution radiation hybrid (RH) mapping, which relies on lines carrying specific radiation-induced chromosomal fragments in nonhomoeologous backgrounds (17) and measures the distance between genetic sites as the frequency with which they remain together after fragmentation. A panel of 184 RH lines developed for chromosome 3B (11) was tested with 65 ISBP markers (table S5) and indicated a resolution of about 263 kb per break. In addition, with a limited set of critical RH lines, we were able to order 35 loci (32 contigs) previously assigned to 3BL7-0.63-1.00 (table S5). We are in the process of increasing the number of RH lines for mapping all the contigs along chromosome 3B.

Further integration of the 3B physical map was achieved through meiotic mapping with a reference genetic map developed from an F_2 population (CsRe) derived from a cross between Chinese Spring (Cs) and the French cultivar Renan (Re). Because the physical contigs originate from Chinese Spring, anchoring them to the CsRe genetic map guarantees high accuracy in ordering. Because BAC libraries are available for both cultivars, this physical map also provides an efficient platform for single-nucleotide polymorphism discovery. To date, 102 SSR and ISBP markers have been mapped with 376 F_2 individuals of the CsRe population (11). Eighty-nine of them were anchored to contigs, permitting the integration of 75 individual contigs (77 Mb) to the genetic map, of which 80% (60/75) were ordered. Using the same criteria as the IBM Neighbors map of maize (18), we also established a 3B neighbor map (11) containing 636 SSR, restriction fragment length polymorphism, sequence tagged site (STS), DArT, and ISBP markers (table S6). In total, 213 contigs were anchored on this map, providing 225 Mb of sequence information for map-based cloning. A Gbrowse interface displaying the integrated chromosome 3B physical map is available at http://urgi.versailles.inra.fr/projects/Triticum/eng/index.php.

We also aligned the wheat 3B physical map against the rice genome using the genic sequences (ESTs/STSs) present on the 75 contigs integrated to the CsRe genetic map (11). Twenty-seven contigs carried 49 ESTs/STSs with homology to 56 orthologous rice genes, including 14 contigs located in the terminal part of the short arm of chromosome 3B, which is collinear with rice chromosome 1S (table S7). In this region, we identified four inversions as well as noncol-

Fig. 1. Relative contribution of ISBP, EST, and SSR markers for anchoring the 3B physical map. Markers found in BAC contigs are indicated in parentheses after the marker type. The numbers of contigs anchored, as well as their physical size (in brackets), are provided for each marker type. The Venn diagram illustrates the relative contribution of each marker type, with the number and total size of contigs anchored by one or more types of markers.

ISBPs (711)
472 contigs
[452 Mb]

296
[221 Mb]

83
[102 Mb]

43
[51 Mb]

87
[64 Mb]

80
[51 Mb]

ESTs (394)
250 contigs
[283 Mb]

SSRs (291)
203 contigs
[219 Mb]

30
[39 Mb]

50
[79 Mb]

Fig. 2. Integrated physical map at the telomeric end of chromosome 3BS and colinearity with rice. (**A**) Genetic map of wheat chromosome 3B. The blue and yellow sectors represent the two distal deletion bins 3BS3-0.87-1.00 and 3BS8-0.78-0.87, respectively. (**B**) BAC contigs integrated to the genetic map. The names and sizes of the 27 contigs are displayed in gray boxes, the sizes of which reflect the relative contig sizes. (**C**) Rice chromosome 1. Each line represents the relationship between an EST located in a wheat BAC contig and the orthologous rice gene. The red sectors indicate rearrangements between the two chromosomes. The four red asterisks designate wheat ESTs for which rice orthologs were found on noncollinear regions or chromosomes other than chromosome 1. The vertical dashed bracket represents a region containing more than 30 kinase-related genes in rice.

linear genes (Fig. 2). This confirms, with a higher degree of resolution, rearrangements observed by deletion mapping between the two most conserved wheat and rice chromosomes (19) and suggests that many local rearrangements have occurred in globally collinear regions since the divergence of wheat and rice more than 50 million years ago. This result also indicates that predicting gene order from sequenced genomes that are not closely related so as to order the physical maps of other genomes requires great caution.

To date, less than a dozen wheat genes have been isolated through map-based cloning. About 40 genes and quantitative trait loci (QTLs) have been identified on chromosome 3B (http://wheat.pw.usda.gov/GG2/maps.shtml) and none have been cloned. Seventeen contigs representing 16.8 Mb are anchored with markers flanking or cosegregating with 16 of these genes and QTLs on our physical map (table S8). With an average contig size of 783 kb, the chromosome 3B physical map allows one to land on any target locus, in a single step, provided that recombination is compatible with fine mapping of the target gene. This has been carried out for the stem rust resistance gene *Sr2* (20) and the QTL *Fhb1* conferring resistance to Fusarium head blight (21) (table S8). The 3B physical map also provides the foundation for sequencing and in-depth studies of the wheat genome composition and organization. Sequencing and annotation are under way for 13 BAC contigs of 800 kb to 3.2 Mb, originating from different regions of chromosome 3B. These large sequenced regions will also aid in SSR and ISBP marker development. Finally, a 15-fold coverage physical map is under way for future chromosome 3B sequencing, which is envisaged for the near future.

By establishing a physical map of the largest wheat chromosome, we demonstrate that the chromosome-based approach is feasible and suitable for the construction of the hexaploid wheat genome physical map. With this physical map, the structure of agronomically important target loci can be defined in a single step and marker development can be accelerated (22), opening up new possibilities for accessing regions important for yield, disease resistance, and trait improvement in wheat. An international collaborative effort is under way now to exploit this new resource and, using the same strategy, projects have begun for the 20 remaining chromosomes (table S1; www.ueb.cas.cz/Olomouc1/LMCC/Resources/resources.html#chrs and www.wheatgenome.org/projects.php), opening the way to future sequencing of the wheat genome.

Finally, although chromosome sorting has so far been applied only to a few other cereals [barley, durum wheat, and rye (8)], important legumes (pea and bean), and trees (Norway spruce) (23), this work exemplifies its broader potential for other complex nonmodel genomes heretofore considered impossible to sequence

despite their socioeconomic importance. Until now, the selection of genomes for sequencing has been determined on the basis of genome simplicity and not agronomic relevance, with serious consequences for crop improvement and food security [for example, by neglecting wheat or choosing the diploid of cotton, *Gossypium raimondii*, to sequence first rather than focusing on the economically important tetraploid *G. hirsutum* (*24*)]. Our work may pave the way for a major change in how the next genomes for de novo sequencing are selected, thereby accelerating improvements in economically important crop species.

References and Notes

1. B. Wollenweber *et al.*, *Curr. Opin. Plant Biol.* **8**, 337 (2005).
2. T. Itoh *et al.*, *Genome Res.* **17**, 175 (2007).
3. R. L. Warren *et al.*, *Genome Res.* **16**, 768 (2006).
4. M. Chen *et al.*, *Plant Cell* **14**, 537 (2002).
5. P. E. Klein *et al.*, *Genome Res.* **10**, 789 (2000).
6. F. Wei *et al.*, *PLoS Genet.* **3**, e123 (2007).
7. M. C. Luo *et al.*, *Genomics* **82**, 378 (2003).
8. J. Dolezel *et al.*, *Chromosome Res.* **15**, 51 (2007).
9. J. Safar *et al.*, *Plant J.* **39**, 960 (2004).
10. T. R. Endo, B. S. Gill, *J. Hered.* **87**, 295 (1996).
11. Materials and methods are available as supporting material on *Science* Online.
12. B. S. Gill *et al.*, *Genome* **34**, 830 (1991).
13. E. Paux *et al.*, *Plant J.* **48**, 463 (2006).
14. J. D. Munkvold *et al.*, *Genetics* **168**, 639 (2004).
15. M. Akbari *et al.*, *Theor. Appl. Genet.* **113**, 1409 (2006).
16. M. Dilbirligi *et al.*, *Genetics* **166**, 461 (2004).
17. K. G. Hossain *et al.*, *Genetics* **168**, 415 (2004).
18. K. C. Cone *et al.*, *Plant Physiol.* **130**, 1598 (2002).
19. M. La Rota, M. E. Sorrells, *Funct. Integr. Genomics* **4**, 34 (2004).
20. R. Kota *et al.*, *Theor. Appl. Genet.* **112**, 492 (2006).
21. S. Liu *et al.*, *Funct. Integr. Genomics* **6**, 83 (2006).
22. M. D. McNeil *et al.*, *Mol. Breed.* **22**, 15 (2008).
23. J. Dolezel *et al.*, *Chromosome Res.* **12**, 77 (2004).
24. Z. J. Chen *et al.*, *Plant Physiol.* **145**, 1303 (2007).
25. We thank D. Boyer, K. Paux, and members of the GENTYANE platform for technical assistance; M. Ganal (Trait Genetics), P. Jack (RAGT), and F. Azanza (Syngenta) for providing SSR markers and genotyping data; the support of F. Legeai and D. Steinbach from Unité de Recherche en Génomique-Info in establishing the Gbrowse display of the integrated physical map; A. Kumar (Choudhury Charan Singh University Meerut), O. Riera-Lizarazu (Oregon State University), and Y. Gu (U.S. Department of Agriculture–Agricultural Research Service) for significant contributions to the RH mapping work; and F. Quetier (Agence Nationale de la Recherche) for support and discussions. Supported by the Agence Nationale de la Recherche (grant ANR-05-BLANC-0258-01) and the INRA.

High-Quality Binary Protein Interaction Map of the Yeast Interactome Network

Haiyuan Yu,[1,2]* Pascal Braun,[1,2]* Muhammed A. Yıldırım,[1,2,3]* Irma Lemmens,[4] Kavitha Venkatesan,[1,2] Julie Sahalie,[1,2] Tomoko Hirozane-Kishikawa,[1,2] Fana Gebreab,[1,2] Na Li,[1,2] Nicolas Simonis,[1,2] Tong Hao,[1,2] Jean-François Rual,[1,2] Amélie Dricot,[1,2] Alexei Vazquez,[5] Ryan R. Murray,[1,2] Christophe Simon,[1,2] Leah Tardivo,[1,2] Stanley Tam,[1,2] Nenad Svrzikapa,[1,2] Changyu Fan,[1,2] Anne-Sophie de Smet,[4] Adriana Motyl,[6] Michael E. Hudson,[6] Juyong Park,[1,7] Xiaofeng Xin,[8] Michael E. Cusick,[1,2] Troy Moore,[9] Charlie Boone,[8] Michael Snyder,[6] Frederick P. Roth,[1,10] Albert-László Barabási,[1,7] Jan Tavernier,[4] David E. Hill,[1,2] Marc Vidal[1,2]†

Current yeast interactome network maps contain several hundred molecular complexes with limited and somewhat controversial representation of direct binary interactions. We carried out a comparative quality assessment of current yeast interactome data sets, demonstrating that high-throughput yeast two-hybrid (Y2H) screening provides high-quality binary interaction information. Because a large fraction of the yeast binary interactome remains to be mapped, we developed an empirically controlled mapping framework to produce a "second-generation" high-quality, high-throughput Y2H data set covering ~20% of all yeast binary interactions. Both Y2H and affinity purification followed by mass spectrometry (AP/MS) data are of equally high quality but of a fundamentally different and complementary nature, resulting in networks with different topological and biological properties. Compared to co-complex interactome models, this binary map is enriched for transient signaling interactions and intercomplex connections with a highly significant clustering between essential proteins. Rather than correlating with essentiality, protein connectivity correlates with genetic pleiotropy.

A crucial step toward understanding cellular systems properties is mapping networks of physical DNA-, RNA-, and protein-protein interactions, the "interactome network," of an organism of interest as completely and accurately as possible. One approach consists in systematically testing all pairwise combinations of predicted proteins to derive the "binary" interactome. Early attempts at binary interactome mapping used high-throughput yeast two-hybrid (Y2H) screening, in which a protein interaction reconstitutes a transcription factor that activates expression of reporter genes. High-throughput Y2H maps have been generated for *Saccharomyces cerevisiae* (*1–3*), *Caenorhabditis elegans* (*4–6*), *Drosophila melanogaster* (*7*), and humans (*8–10*). An alternative approach consists in generating "co-complex" interactome maps, achievable by high-throughput coaffinity purification followed by mass spectrometry (AP/MS) to identify proteins bound to tagged baits, as done for *Esche-*richia coli (*11*, *12*), *S. cerevisiae* (*13–16*), and humans (*17*).

To investigate fundamental questions of interactome network structure and function, it is necessary to understand how the size and quality of currently available maps, including thorough evaluation of differences between binary and co-complex maps, might have affected conclusions about global and local properties of interactome networks (*18*, *19*). Here, we address these issues using the yeast *S. cerevisiae* as a model system.

First, we compared the quality of existing high-throughput binary and co-complex data sets to information obtained from curating low-throughput experiments described in the literature (Fig. 1A). For binary interactions, we examined (i) the subset found by Uetz *et al.* in a proteome-scale all-by-all screen ("Uetz-screen"), excluding the pairs found in a focused, potentially biased experiment involving only 193 baits ("Uetz-array") (*2*); and (ii) the Ito *et al.* interactions found three times or more ("Ito-core"), independently from those found one or two times ("Ito-noncore"), a distinction recommended by the authors but seldom applied in the literature (*3*). For co-complex associations, we investigated two high-throughput AP/MS data sets referred to as "Gavin" (*15*) and "Krogan" (*16*). For literature-curated interactions, we considered only those curated from two or more publications ("LC-multiple") (*20*), which we judged of higher quality than those curated from a single publication.

To experimentally compare the quality of these data sets, we selected a representative sample of ~200 protein interaction pairs from each one and tested them by means of two independent interaction assays, Y2H and a yellow fluorescent protein complementation assay (PCA) (*21*) [Supporting Online Material (SOM) I]. In PCA, bait and prey proteins are fused to nonfluorescent fragments of yellow fluorescent protein that, when brought in close proximity by interacting proteins, reconstitute a fluorescent protein in mammalian cells. In con-