nature genetics
OPEN

Hybridization of powdery mildew strains gives rise to pathogens on novel agricultural crop species

Fabrizio Menardo[1], Coraline R Praz[1], Stefan Wyder[1], Roi Ben-David[1,4], Salim Bourras[1], Hiromi Matsumae[2], Kaitlin E McNally[1], Francis Parlange[1], Andrea Riba[3], Stefan Roffler[1], Luisa K Schaefer[1], Kentaro K Shimizu[2], Luca Valenti[1], Helen Zbinden[1], Thomas Wicker[1,5] & Beat Keller[1,5]

Throughout the history of agriculture, many new crop species (polyploids or artificial hybrids) have been introduced to diversify products or to increase yield. However, little is known about how these new crops influence the evolution of new pathogens and diseases. Triticale is an artificial hybrid of wheat and rye, and it was resistant to the fungal pathogen powdery mildew (*Blumeria graminis*) until 2001 (refs. 1–3). We sequenced and compared the genomes of 46 powdery mildew isolates covering several *formae speciales*. We found that *B. graminis* f. sp. *triticale*, which grows on triticale and wheat, is a hybrid between wheat powdery mildew (*B. graminis* f. sp. *tritici*) and mildew specialized on rye (*B. graminis* f. sp. *secalis*). Our data show that the hybrid of the two mildews specialized on two different hosts can infect the hybrid plant species originating from those two hosts. We conclude that hybridization between mildews specialized on different species is a mechanism of adaptation to new crops introduced by agriculture.

© 2016 Nature America, Inc. All rights reserved.

The artificial hybrid triticale was introduced into commercial agriculture in the 1960s. The hexaploid triticale genome is composed of genomes A and B from wheat and the rye genome, R (AABBRR)[3]. Triticale was initially resistant to powdery mildew; however, this pathogen was first observed on triticale in 2001 and has since become a major disease in Europe[1,2]. We sequenced 46 isolates of *B. graminis* (including the 96224 reference isolate[4]) from different European countries and Israel (**Supplementary Table 1**) with host ranges corresponding to four different *formae speciales*, including the previously described rye (*B.g. secalis*) and wheat (*B.g. tritici*) powdery mildews. Our infection tests showed that *B.g. secalis* grows exclusively on rye, whereas *B.g. tritici* is able to grow on tetraploid (durum) and hexaploid (bread) wheat (**Table 1**, **Supplementary Figs. 1–23**, **Supplementary Table 2** and **Supplementary Note**). On the basis of infection tests, we define the *triticale forma specialis* as being able to grow on triticale, on hexaploid and tetraploid wheat (with lower penetration efficiency; **Supplementary Fig. 24** and **Supplementary Note**), and to a very limited extent on rye. Furthermore, we designate mildew growing exclusively on tetraploid wheat as new *forma specialis*, *B. graminis* f. sp. *dicocci*[5].

Overall, the genomes of the 46 isolates were very similar to each other, allowing high-quality mapping of the 45 resequenced isolates to the 96224 reference isolate (**Supplementary Note**). We identified between 115,543 and 332,450 polymorphic nucleotide sites per isolate in comparison to the *B.g. tritici* reference genome (**Supplementary Figs. 25** and **26**, and **Supplementary Note**). Principal-component analysis (PCA) based on 717,701 polymorphic sites clearly distinguished four groups (**Fig. 1**) that corresponded to the four *formae speciales* identified in the infection tests. This grouping indicates that gene flow is restricted between *formae speciales*. Interestingly, the *B.g. triticale* isolates formed a group distinct from the *B.g. tritici* isolates. This contradicts the current hypothesis that the host range of *B.g. tritici* expanded to include triticale through mutation of a few genes[2,6,7]. Instead, our analysis shows that *B.g. triticale* isolates form a specific group with a distinct evolutionary history (**Supplementary Figs. 27–30** and **Supplementary Note**).

From genome comparisons, we identified sets of polymorphisms that were fixed in the four *formae speciales* (polymorphisms shared by all isolates of a particular *forma specialis*). Such fixed polymorphisms between *formae speciales* (substitutions) will hereafter be referred to as the 'genotype' of a *forma specialis*. Interestingly, the *B.g. triticale* genome consisted of large genomic segments that had the same genotype as *B.g. secalis* alternating with segments with a *B.g. tritici* genotype (**Fig. 2a**,**b** and **Supplementary Figs. 31–33**). Such patterns are characteristic of recent hybrids, reflecting a limited number of recombination events between parental genotypes. Furthermore, analysis of nucleotide diversity in the genomes of the different *formae speciales* showed that *B.g. triticale* had a characteristic distribution with multiple peaks that is consistent with the hypothesis of a recent origin by hybridization (**Fig. 2c**,**d** and **Supplementary Note**). We tested this hypothesis on the 172,274 fixed polymorphisms that distinguish *B.g. tritici* from *B.g. secalis*. We found that, in *B.g. triticale*, between 11.9 and 21.4% of the polymorphic sites represented the *B.g. secalis* genotype. These polymorphic sites were

[1]Department of Plant and Microbial Biology, University of Zürich, Zurich, Switzerland. [2]Institute of Evolutionary Biology and Environmental Studies, University of Zürich, Zurich, Switzerland. [3]Biozentrum, University of Basel, Basel, Switzerland. [4]Present address: Institute of Plant Science, ARO-Volcani Center, Bet Dagan, Israel. [5]These authors jointly supervised this work. Correspondence should be addressed to T.W. (wicker@botinst.uzh.ch) or B.K. (bkeller@botinst.uzh.ch).

Received 17 July 2015; accepted 11 December 2015; published online 11 January 2016; doi:10.1038/ng.3485

Table 1 Host specificity of powdery mildew *formae speciales*

	Tetraploid wheat	Hexaploid wheat	Rye	Triticale
B.g. dicocci	+	–	–	–
B.g. tritici	+	+	–	–
B.g. secalis	–	–	+	–/(+)
B.g. triticale	+	+	+/–	+

Columns represent different host species, and rows represent different *formae speciales* of *B. graminis*. A compatible interaction is indicated by a plus sign, whereas an incompatible interaction is indicated by a minus sign. Intermediate phenotypes are indicated by "+/–", referring to *B.g. triticale* isolates that show limited growth on rye. "–/(+)" indicates the very reduced level of growth that we observed for some *B.g. secalis* isolates on triticale.

located on genomic segments that made up between 6.6 and 17.3% of the genomes of the *B.g. triticale* isolates. In contrast, over 80% of these genomes had the *B.g. tritici* genotype. Thus, we conclude that *B.g. triticale* is a hybrid of *B.g. tritici* and *B.g. secalis* and that the hybridization happened very recently (**Supplementary Table 3** and **Supplementary Note**).

Furthermore, phylogeographic analysis and the proportions of the parental genomes suggest that the initial hybridization event was followed by two backcrosses with *B.g. tritici* and that this process likely occurred in Europe (**Supplementary Figs. 34–37** and **Supplementary Note**).

These findings raised the question of whether *B.g. triticale* originated from a single hybridization of *B.g. secalis* and *B.g. tritici* (followed by two backcrosses) or whether more individuals from the two *formae speciales* also contributed. If the first hybridization was a single event that involved only one *B.g. secalis* isolate, we would expect to observe no diversity in the *B.g. triticale* portion of the genome inherited from *B.g. secalis*. On the basis of phylogenetic analysis, we found 66 genes for which all *B.g. triticale* isolates inherited the *B.g. secalis* gene. We counted the number of haplotypes present in *B.g. triticale* for each of these 66 genes and found six genes that showed two different haplotypes. These different haplotypes were also present in the five sequenced *B.g. secalis* isolates. We conclude that the minimum number of *B.g. secalis* individuals that contributed to *B.g. triticale* is two, which is likely the result of two independent origins of *B.g. triticale* (**Supplementary Fig. 38** and **Supplementary Note**).

B. graminis, as with most fungi, has two mating types[4]. Because the mating types of *B.g. secalis* and *B.g. tritici* have different genotypes, we could trace back the origin of mating types in *B.g. triticale*. We found that all *B.g. tritici* partners in the first hybridizations seemed to have been of the MAT1-1-3 mating type whereas the *B.g. secalis* partners were of the MAT1-2-1 mating type. The *MAT1-2-1* locus of the *B.g. tritici* genotype was acquired by *B.g. triticale* through one of the backcrosses, whereas the *MAT1-1-3* locus of the *B.g. secalis* genotype was not present in any of the sequenced *B.g. triticale* isolates (**Supplementary Figs. 39** and **40**, and **Supplementary Note**). These findings are in contrast to those on the plant pathogen *Zymoseptoria pseudotritici*, whose origin was traced back to a single hybridization event[8]. Unlike *Z. pseudotritici*, the new *forma specialis B.g. triticale* likely did not pass through a bottleneck, as the multiple isolates involved in hybridization passed on a considerable proportion of parental genetic diversity (**Supplementary Table 4** and **Supplementary Note**).

On the basis of the number of observed recombination events between the *B.g. tritici* and *B.g. secalis* genotypes, we estimated that *B.g. triticale* isolates underwent between seven and 47 (depending on the isolate) sexual cycles from the origin of the *forma specialis* (**Supplementary Fig. 41** and **Supplementary Note**). As *B. graminis* has a maximum of one sexual cycle per year[4], we conclude that *B.g. triticale* originated after the introduction of triticale as a commercial crop in the 1960s, possibly multiple times, and since then different isolates have undergone a different number of sexual generations. Because *B.g. triticale* is a very recent hybrid, we studied the phylogenetic relationship of its two parents. We estimated the divergence time for the parental *formae speciales* of *B.g. triticale*—*B.g. secalis* and *B.g. tritici*—using 206 orthologous single-copy genes and found that they diverged between 168,245 and 240,169 years ago (**Supplementary Fig. 42**, **Supplementary Table 5** and **Supplementary Note**). In contrast, their hosts—rye and wheat, respectively—had already diverged approximately 4 million years ago[9]. This incongruence in the divergence times of a host and pathogen can be due to host tracking: until a few hundreds of thousands of years ago, wheat and rye were still in the host range of a single *forma specialis*, and divergence into two distinct *formae speciales* occurred only relatively recently. This history would be consistent with previous findings that the *formae speciales B.g. tritici* and *B. graminis* f. sp. *hordei* diverged approximately 6 million years ago[4], whereas their hosts, wheat and barley, respectively, diverged at least 2 million years earlier[9]. An alternative explanation could be a recent host jump from wheat to rye or vice versa, followed by the rapid emergence of barriers to gene flow.

Despite our large sequencing effort, the molecular basis for the host range expansion of *B.g. triticale* remains obscure. Genetic determinants of the expansion of host specificity to triticale are likely located on genomic segments that were inherited from *B.g. secalis*. Among these genes might be the genetic determinant for the host range expansion of *B.g. triticale*. Six of the 66 *B.g. secalis* genes inherited by all *B.g. triticale* isolates encoded putative effectors (proteins that are secreted into the host cell to facilitate pathogen proliferation) (**Supplementary Table 6** and **Supplementary Note**). In *B. graminis*, effectors are typically small proteins with short, characteristic motifs that can be identified bioinformatically[4] (through the presence of signal peptide and lack of homology with the protein domains of other organisms). However, transcriptome profiling of *B.g. triticale* grown

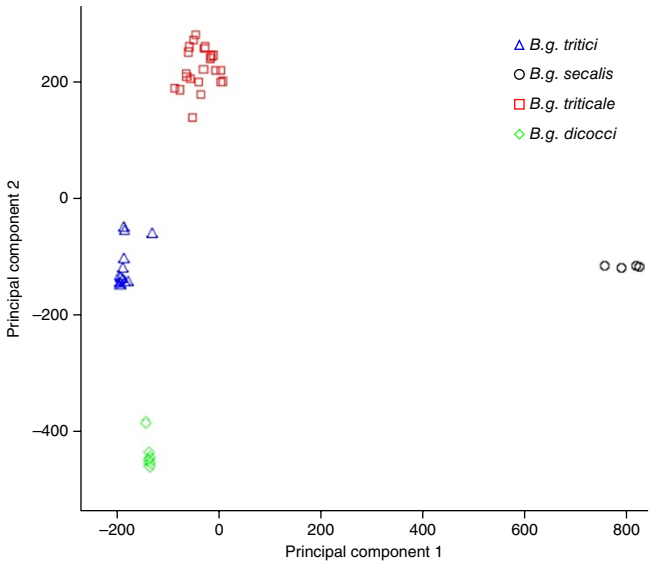

Figure 1 Analysis of genome sequence diversity in 46 *B. graminis* isolates. PCA was based on 717,701 SNPs that are polymorphic in at least two isolates. The PCA differentiates four groups corresponding to the four different *formae speciales*, which are defined by their different host ranges.

© 2016 Nature America, Inc. All rights reserved.

on wheat and triticale showed that none of these six genes were differentially expressed by the fungus in the two different hosts. In general, the transcriptome profile of *B.g. triticale* was mostly independent of the host species (**Supplementary Figs. 43–47** and **Supplementary Note**). Alternatively, host specificity could be a quantitative trait that requires a certain number of genes with partially overlapping and/or complementary functions. Effector genes in particular are thought to have overlapping or partially redundant functions[4,10].

It is intriguing that the host of *B.g. triticale* (triticale) is a hybrid of rye and bread wheat, which are the hosts of the ancestors of *B.g. triticale*. Interestingly, bread wheat (the host of *B.g. tritici*) is itself the result of a hybridization that occurred approximately 10,000 years ago between domesticated tetraploid (emmer) wheat and the diploid wild grass *Aegilops tauschii*[11]. Moreover, the host range of *B.g. tritici* (hexaploid and tetraploid wheat) encompasses the host range of *B.g. dicocci* (tetraploid wheat). This is reminiscent of the expanded host range of

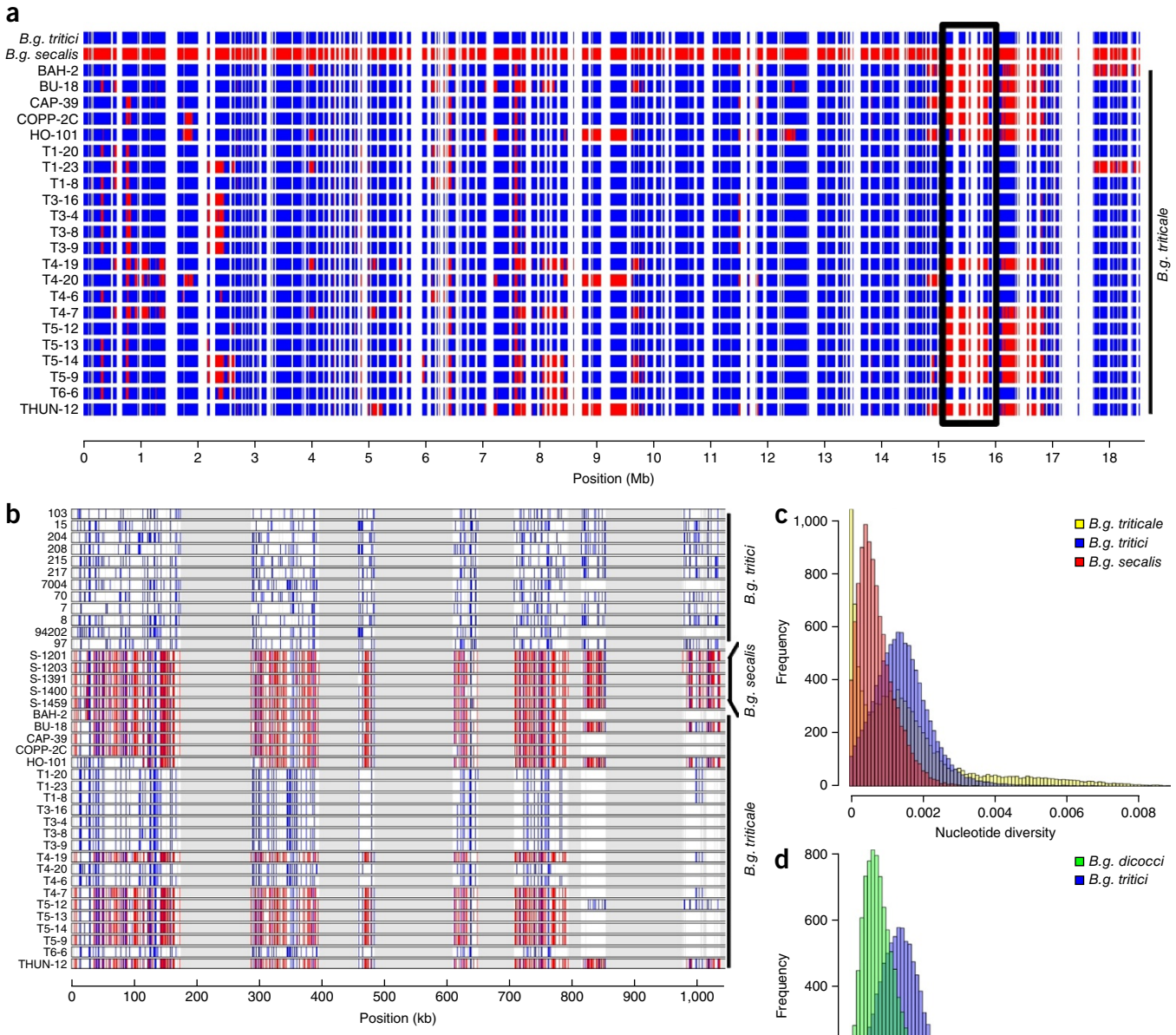

Figure 2 Examples of nucleotide diversity patterns in powdery mildew isolates. (**a**) Nucleotide substitutions in *B.g. triticale* isolates in comparison to *B.g. tritici* and *B.g. secalis* on *B. graminis* linkage group 3 (ref. 28). Fixed polymorphisms specific for *B.g. tritici* isolates are depicted in blue, whereas those specific for *B.g. secalis* isolates are depicted in red. The black box shows the section of the linkage group represented in **b**. (**b**) Polymorphism patterns on physical contig 140 of linkage group 3 in *B.g. tritici* (top 12 lines), *B.g. secalis* (following five lines) and *B.g. triticale* (last 22 lines) isolates. Polymorphisms as compared to the *B.g. tritici* reference genome sequence (96224) that are present in at least one of the *B.g. tritici* isolates are colored in blue in all isolates of all *formae speciales*. Polymorphisms not present in *B.g. tritici* but in *B.g. secalis* are colored in red in *B.g. secalis* and *B.g. triticale* isolates. Gray windows represent sequence gaps in the *B.g. tritici* reference genome, and white areas represent non-polymorphic regions. (**c**) Histograms of the nucleotide diversity (π) of aligned genome windows larger than 10 kb in size between *B.g. triticale* (yellow), *B.g. tritici* (blue) and *B.g. secalis* (red) isolates (**Supplementary Note**). (**d**) Histograms of the nucleotide diversity (π) of aligned genome windows (larger than 10 kb in size) in *B.g. tritici* (blue) and *B.g. dicocci* (green) isolates. The distribution of *B.g. tritici* is shifted to the right as compared to the distribution of *B.g. dicocci*, indicating greater diversity in *B.g. tritici* than in *B.g. dicocci*.

© 2016 Nature America, Inc. All rights reserved.

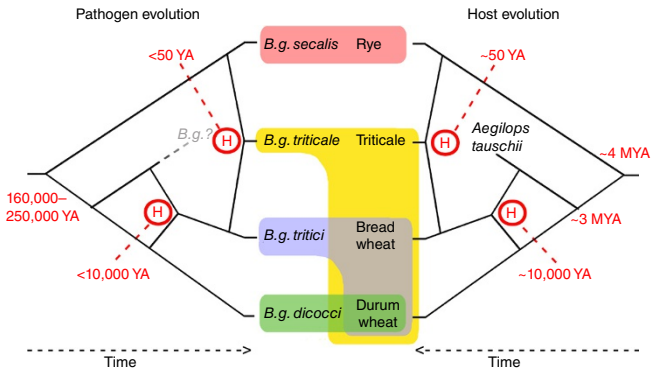

Figure 3 Model for the evolution of specialized forms and host ranges in *B. graminis*. Phylogenetic trees of *B. graminis* and its hosts are represented facing each other, indicating that pathogen evolution mirrors the evolution of the host species. The branch corresponding to the unknown *forma specialis* (*B.g.?*) that hybridized with *B.g. dicocci* to form *B.g. tritici* is shown in gray. Hybridization events are marked with "H". Estimated times of species and *forma specialis* divergence and hybridization events are shown in red (YA, years ago; MYA, million years ago). The origin times of *B.g. triticale* and *B.g. tritici* are based on the origin times of the host species. The host ranges of the different *formae speciales* are indicated as shaded areas of different color.

B.g. triticale that encompasses the host range of one of its parent species, *B.g. tritici* (wheat), in addition to triticale. We therefore hypothesized that *B.g. tritici* could also be a hybrid between a pathogen of tetraploid wheat (*B.g. dicocci*) and one that infects *A. tauschii*. However, in contrast to *B.g. triticale* in which we found genomic segments of the *B.g. tritici* genotype, we did not find large genomic segments in *B.g. tritici* isolates that could be assigned to a *B.g. dicocci* genotype. This absence could be explained by multiple rounds of recombination that have eroded the characteristic pattern of sequence segments. A different approach to test hybridization that is more resistant to the action of recombination makes use of gene genealogies[12], which have been used to identify hybridization in different organisms[13,14]. We used coalescent-based methods to calculate the probability of the evolutionary model in which both *B.g. tritici* and *B.g. triticale* are hybrids and four alternative models in which there is no hybridization or only one of the two *formae speciales* is a hybrid[15–18]. The model depicted in **Figure 3** with two hybridizations seems to be the most likely (**Supplementary Fig. 48**, **Supplementary Tables 7** and **8**, and **Supplementary Note**). However, we cannot completely rule out the possibility that *B.g. tritici* originated through a more complex untested scenario with multiple past admixture events.

We conclude that *B.g. tritici* arose sometime after the emergence of bread wheat, probably several thousand years ago, through hybridization of a *B.g. dicocci* strain and a different, yet unknown mildew strain. Hybridization has been reported to be important for adaptation to new hosts in several fungal and oomycete pathogens[19–24]. Our data now indicate that hybridization was the causal step in host range expansion to a newly bred or evolved species. It is apparent that co-evolution based on hybridization is a likely evolutionary pathway for *B. graminis* that infects wheat and other grasses, which themselves evolve predominantly through hybridization[25]. It is particularly fascinating that pathogen evolution mirrors evolution on the host side and that the hybrid of two mildews specialized on two different hosts can infect the hybrid plant species originating from those two hosts (**Fig. 3**). It is possible that our findings define a more general evolutionary pattern: stem rust (*Puccinia graminis*) was reported to infect triticale at approximately the same time as *B. graminis*[26]. Also, in *P. graminis*, hybridization between *formae speciales* has been reported to be important for host range determination[27]. It is noteworthy that this pattern of evolution is observed in pathogens of agricultural species. It is possible that agricultural ecosystems increase the possibility that different pathogens co-occur in large populations, thus making hybridization more likely than in natural ecosystems. Many agricultural crops are polyploids and/or contain resistance genes introgressed from wild relatives. Little is known about what determines the durability of such introgressed genes, but hybridization of pathogens of the wild relative with the crop pathogen should be considered in resistance breeding approaches. In addition, in the future production of man-made crops, the genetic resistance of parents should be carefully selected to make resulting hybrids more durably resistant in the field.

URLs. *B.g.* f. sp. *hordei* genome (BluGen), http://www.blugen.org/; *Neurospora crassa* genome (Broad Institute), http://www.broadinstitute.org/annotation/genome/neurospora/; Triticeae transposable elements (PTREP12) databases, http://www.botinst.uzh.ch/en/research/genetics/thomasWicker/TREP.html.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. All genomic and transcriptomic sequences used for this study can be accessed at the Sequence Read Archive (SRA) under accession SRP062198 and at the Gene Expression Omnibus (GEO) under accessions GSE73399, respectively.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We want to thank S. Brunner for advice and support, A. Dinoor, V. Troch and F. Mascher for mildew isolates, C. Aquino and H. Rehrauer for their help with sequencing and bioinformatic analysis, and A. Widmer for reading and commenting on the manuscript. This work was supported by the University Research Priority Program (URPP) 'Evolution in Action' of the University of Zurich, Swiss National Science Foundation grant 310030B_144081/1 , a grant from the Swiss Federal Office for Agriculture and an Advanced Investigator Grant from the European Research Council (ERC-2009-AdG 249996, Durableresistance).

AUTHOR CONTRIBUTIONS

T.W. and B.K. designed the project and wrote the manuscript. T.W., S.R. and F.M. wrote software for the analysis. A.R. and F.M. performed statistical analysis. R.B.-D., C.R.P., L.V., K.E.M. and F.M. phenotyped and propagated the isolates. R.B.-D. and F.M. collected isolates and extracted DNA for sequencing. F.P. and F.M. performed crosses between mildew isolates. H.Z. performed RNA extraction. K.K.S. discussed the population genetics analysis. L.K.S. performed staining of infected leaves. S.B. developed the staining protocol. C.R.P. and F.M. analyzed RNA-seq data. S.W., H.M., T.W., S.R., C.R.P. and F.M. performed bioinformatics and population genetics analyses.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

This work is licensed under a Creative Commons Attribution- NonCommercial-ShareAlike 3.0 Unported License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/.

1. Mascher, F., Reichmann, P. & Schori, A. Impact de l'oïdium sur la culture du triticale. *Revue Suisse Agric.* **38**, 193–196 (2005).

© 2016 Nature America, Inc. All rights reserved.

2. Walker, A.S., Bouguennec, A., Confais, J., Morgant, G. & Leroux, P. Evidence of host-range expansion from new powdery mildew (*Blumeria graminis*) infections of triticale (×*Triticosecale*) in France. *Plant Pathol.* **60**, 207–220 (2011).
3. Oettler, G. The fortune of a botanical curiosity—triticale: past, present and future. *J. Agric. Sci.* **143**, 329–346 (2005).
4. Wicker, T. *et al.* The wheat powdery mildew genome shows the unique evolution of an obligate biotroph. *Nat. Genet.* **45**, 1092–1096 (2013).
5. Eshed, N., Dinoor, A. & Litwin, Y. The physiological specialization of wheat powdery mildew in Israel and the search for mildew resistance in wild wheat *Triticum dicoccoides*. *Phytoparasitica* **22**, 49–90 (1994).
6. Troch, V., Audenaert, K., Bekaert, B., Höfte, M. & Haesaert, G. Phylogeography and virulence structure of the powdery mildew population on its 'new' host triticale. *BMC Evol. Biol.* **12**, 76 (2012).
7. Troch, V. *et al.* Evaluation of resistance to powdery mildew in triticale seedlings and adult plants. *Plant Dis.* **97**, 410–417 (2013).
8. Stukenbrock, E.H., Christiansen, F.B., Hansen, T.T., Dutheil, J.Y. & Schierup, M.H. Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. *Proc. Natl. Acad. Sci. USA* **109**, 10954–10959 (2012).
9. Middleton, C.P. *et al.* Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the Triticeae tribe. *PLoS One* **9**, e85761 (2014).
10. Birch, P.R.J. *et al.* Oomycete RXLR effectors: delivery, functional redundancy and durable disease resistance. *Curr. Opin. Plant Biol.* **11**, 373–379 (2008).
11. Salamini, F., Ozkan, H., Brandolini, A., Schäfer-Pregl, R. & Martin, W. Genetics and geography of wild cereal domestication in the near east. *Nat. Rev. Genet.* **3**, 429–441 (2002).
12. Hein, H., Schierup, M.H. & Wiuf, C. *Genes Genealogies, Variation and Evolution. A Primer in Coalescent Theory* (Oxford University Press, 2005).
13. Xu, J., Vilgalys, R. & Mitchell, T.G. Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*. *Mol. Ecol.* **9**, 1471–1481 (2000).
14. Sota, T. Radiation and reticulation: extensive introgressive hybridization in the carabid beetles *Ohomopterus* inferred from mitochondrial gene genealogy. *Popul. Ecol.* **44**, 145–156 (2002).
15. Degnan, J.H. & Salter, L.A. Gene tree distributions under the coalescent process. *Evolution* **59**, 24–37 (2005).
16. Than, C., Ruths, D. & Nakhleh, L. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* **9**, 322 (2008).
17. Yu, Y., Than, C., Degnan, J.H. & Nakhleh, L. Coalescent histories on phylogenetic networks and detection of hybridization despite incomplete lineage sorting. *Syst. Biol.* **60**, 138–149 (2011).
18. Yu, Y., Degnan, J.H. & Nakhleh, L. The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genet.* **8**, e1002660 (2012).
19. Brasier, C.M., Cooke, D.E.L. & Duncan, J.M. Origin of a new *Phytophthora* pathogen through interspecific hybridization. *Proc. Natl. Acad. Sci. USA* **96**, 5878–5883 (1999).
20. Brasier, C. The rise of the hybrid fungi. *Nature* **405**, 134–135 (2000).
21. Newcombe, G., Stirling, B., McDonald, S. & Bradshaw, H.D. *Melampsora* × *columbiana*, a natural hybrid of *M. medusae* and *M. occidentalis*. *Mycol. Res.* **104**, 261–274 (2000).
22. Goss, E.M. *et al.* The plant pathogen *Phytophthora andina* emerged via hybridization of an unknown *Phytophthora* species and the Irish potato famine pathogen, *P. infestans*. *PLoS One* **6**, e24543 (2011).
23. Brasier, C.M. & Kirk, S.A. Rapid emergence of hybrids between the two subspecies of *Ophiostoma novo-ulmi* with a high level of pathogenic fitness. *Plant Pathol.* **59**, 186–199 (2010).
24. Farrer, R.A. *et al.* Multiple emergences of genetically diverse amphibian-infecting chytrids include a globalized hypervirulent recombinant lineage. *Proc. Natl. Acad. Sci. USA* **108**, 18732–18736 (2011).
25. Feldman, M. & Levy, A.A. Genome evolution due to allopolyploidization in wheat. *Genetics* **192**, 763–774 (2012).
26. Tian, S., Weinert, J. & Wolf, G.A. Infection of triticale cultivars by *Puccinia striiformis*: first report on disease severity and yield loss. *J. Plant Dis. Protection* **111**, 461–464 (2004).
27. Luig, N.H. & Watson, I.A. The role of wild and cultivated grasses in the hybridization of formae speciales of *Puccinia graminis*. *Aust. J. Biol. Sci.* **25**, 335–342 (1972).

© 2016 Nature America, Inc. All rights reserved.

ONLINE METHODS

Host specificity tests. To test the host specificity of the isolates used in this work, we infected six cultivars of triticale, five cultivars of wheat (three hexaploid and two tetraploid) and three cultivars of rye with 46 *Blumeria* isolates (**Supplementary Table 2**). The plant cultivars used in the host specificity tests were chosen for the absence of known race-specific resistance genes and known high susceptibility to many tested mildew isolates. In this way, we avoided race-specific resistance as a confounding factor in the determination of host specificity. We infected 10-d-old detached leaf segments with fresh spores, and the infected leaf segments were kept on benzimidazole agar plates at 20 °C and 70% humidity with a 16-h light/8-h dark cycle. We scored the phenotypes 10 d after infection using three categories: virulent, intermediate and avirulent (**Supplementary Figs. 1–23** and **Supplementary Note**).

Staining for assessment of penetration efficiency. Infected leaf segments were collected 2 d after infection and incubated for 4 d in destaining solution (8.3% lactic acid and 16.7% glycerol in ethanol). Leaves were stained for 45 s with neutral red (0.1%), which was used as a contrasting agent to observe haustoria inside plant epidermal cells. Aerial fungal structures were stained for 45 s using 0.25% Coomassie blue.

Sequencing, mapping, SNP calling, assembly and principal-component analysis. DNA extraction was performed following the methods of Bourras and colleagues[28]. 101-bp paired-end libraries were created and sequenced with the Illumina HiSeq platform at the Functional Genomic Center of Zurich or at the Genome Analysis Centre at Norwich Research Park (TGAC), obtaining between 31,039,416 and 62,834,815 reads (mean of 41,810,764 reads) (**Supplementary Table 1**). Reads were mapped on the *B.g. tritici* reference genome[4] using Bowtie 2.1.0 (ref. 29) with option --score-min L, -0.6, -0.25. We used the following command in SAMtools 0.1.19 (ref. 30) to convert formats and collect information about single genomic positions: view, sort, mpileup -q 15 (only reads with mapping quality greater than 15 were considered). Finally, we used bcftools to generate a VCF file that was parsed with in-house Perl scripts (available upon request). We considered as high-confidence SNPs only positions with a minimum mapping score of 20, a minimum coverage of 20× and a minimum frequency of the alternative call of 0.95. Polymorphisms were considered to be fixed if they were found to be present in all isolates of a *forma specialis* and were different in the other *formae speciales*.

De novo assembly of the genomes for all isolates was performed with CLC Genomic Workbench 7.5 using standard parameters. PCA was performed with the R package GAPIT[31,32].

Identification of orthologous genes. We used GMAP (version 2013-07-20)[33] to annotate genes in the assembly of isolate S-1201 (*B.g. secalis* reference genome), using the 7,186 genes of the *B.g. tritici* reference genome (isolate 96224) as a template. We found 6,864 genes in the genome assembly of isolate S-1201. The protein and CDS databases for *B.g. hordei* and *Neurospora crassa* were downloaded from BluGen and the Broad Institute, respectively (accessed 1 March 2014; see URLs). After the elimination of genes with homology to the *Blumeria* repeat database (BRD) and the Triticeae transposable elements (PTREP12) databases, we retained 9,733 genes for *N. crassa* and 6,011 genes for *B.g. hordei*. To cluster genes in families, we used all-against-all BLAST searches[34] and grouped together genes that were reciprocal hits with a minimum alignment length of 150 bp and a BLASTN *e* value $\leq 1 \times 10^{-10}$. We then defined as single-copy gene families all families with four genes, one for each of the species used (*N. crassa*, *B.g. tritici*, *B.g. hordei* and *B.g. secalis*), that were reciprocal top BLAST hits. With these criteria, we obtained 208 orthologous gene families. We used GMAP (version 2013-07-20)[33] to retrieve these genes in all the *B.g. tritici* and *B.g. secalis* isolates, using as a template the genes from the same *forma specialis* (isolates 96224 and S-1201, respectively). Assemblies for the *B.g. hordei* isolates AOIY01 and AOLT01 were obtained from Hacquard and colleagues[35]. We annotated the 208 genes on these two assemblies with GMAP (version 2013-07-20)[33] using the genes of the *B.g. hordei* reference isolate as the template. We could not find two genes in all *B.g. hordei* isolates, and we therefore excluded them. The final data set was composed of 206 orthologous single-copy genes from *N. crassa*, three from *B.g. hordei* isolates, five from *B.g. secalis* isolates and 13 from *B.g. tritici* isolates. This data set was used to infer a species tree phylogeny and to estimate divergence time (**Supplementary Note**).

We used analogous methods to define single-copy orthologous genes in *B. graminis* (without *N. crassa*). We used all-against-all BLAST searches (between *B.g. hordei* (reference isolate), *B.g. tritici* (96224), *B.g. secalis* (S-1201) and *B.g. dicocci* (220)). We grouped together genes that were reciprocal BLAST hits with a minimum alignment length of 150 bp and a BLASTN *e* value $\leq 1 \times 10^{-10}$. We then defined as single-copy gene families all families with four genes, one for each of the *formae speciales* used (*B.g. dicocci*, *B.g. tritici*, *B.g. hordei* and *B.g. secalis*), that were top reciprocal hits. This resulted in 4,556 single-copy genes. We then retrieved these genes in all isolates, using as a template the genes from the same *forma specialis* (96224 for *B.g. tritici*, S-1201 for *B.g. secalis* and 220 for *B.g. dicocci*); the reference isolate of *B.g. hordei* was used as the outgroup. Multiple-sequence alignments were generated with Muscle[36], and we inferred maximum-likelihood trees for all alignments with RAxML 8.0.22 (ref. 37) using a generalized time-reversible (GTR) + Gamma model[38–40]. We used Newick Tools[41] to identify particular topology patterns. This data set was used in phylogeographic analysis, in the identification of genes inherited from *B.g. secalis* in all *B.g. triticale* isolates, in the estimation of the minimum number of isolates that contributed to the first hybridization and in the coalescent-based estimation of the most likely evolutionary network (**Supplementary Note**).

Alignment and phylogeny for divergence time estimation. Protein multiple-sequence alignments were generated with Muscle 3.8.31 (ref. 36) and reverse translated into nucleotide sequences with TranslatorX v1.1 (ref. 42). The concatemer of all the alignments was 410,189 bp long. Phylogenetic inference on the partitioned data set was performed with MrBayes 3.2.2 (ref. 43). We ran two independent replications of 10,000,000 generations. Variation in substitution rates across sites was modeled with a discretized (four categories) gamma distribution[39,40]. The chains have been let free to sample all models of the GTR model family using reversible-jump Monte Carlo Markov chain[44]. The divergence time between *B.g. hordei* and *B.g. tritici* was used as the calibration point (5.2–7.4 million years ago[4] or 10,000–14,000 years ago[45]) (**Supplementary Note**) under the independent gamma rate relaxed-clock model (white noise model)[46].

RNA extraction and transcriptome analysis. Leaf segments of the wheat variety Chinese Spring and the triticale variety Timbo were infected with two different *B.g. triticale* isolates (THUN-12 and T3-8). Leaves were left on benzimidazole agar plates as described in Parlange *et al.*[47] and collected 2 d after infection. Each pathogen-host combination was replicated three times. RNA extraction was performed with the miRNeasy mini kit from Qiagen according to the manufacturer's recommendations. 125-bp single-end libraries were created and sequenced with the Illumina HiSeq platform at the Functional Genomics Center Zurich. RNA sequencing (RNA-seq) reads were mapped with STAR[48] (allowing four mismatches per 100 bp). Read counts were determined with featureCounts 1.4.6 (ref. 49). The R package edgeR was used for statistical analysis, and genes were tested for differential expression with a generalized linear model and tagwise estimation of dispersion[50].

Testing of evolutionary networks with PhyloNet. The probabilities of five different evolutionary hypotheses (networks a–e; graphically represented in **Supplementary Fig. 35**) were evaluated with a coalescent-based method implemented in PhyloNet[16] (**Supplementary Note**). Because of computational limitations, we generated gene trees using two subsets of isolates. We tested two different, non-overlapping sets of isolates (one *B.g. hordei*, two *B.g. secalis* and three each of *B.g. tritici*, *B.g. dicocci* and *B.g. triticale*, for a total of 12 isolates for each data set), each of them with ten different sets of 300 randomly selected single-copy genes for which trees were inferred with RAxML[37]. The composition of the first data set was as follows: *B.g. hordei* reference isolate; 96224, 97 and 103 (*B.g. tritici*); S-1201 and S-1400 (*B.g. secalis*); THUN-12, T3-8 and COPP-2C (*B.g. triticale*); and 58, 66 and 220 (*B.g. dicocci*). The composition of the second data set was the following: *B.g. hordei* reference isolate; 94202, 8 and 70 (*B.g. tritici*); S-1203 and S-1459 (*B.g. secalis*); T1-23, T6-6 and HO-101 (*B.g. triticale*); and 63, 207 and 209 (*B.g. dicocci*). To compare the likelihoods

© 2016 Nature America, Inc. All rights reserved.

doi:10.1038/ng.3485

of models with a different number of parameters, we used three information criteria—the Akaike information criterion[51] (AIC), the corrected AIC (AICc) and the Bayesian information criterion[52] (BIC). These measures give increasing penalties to models containing more parameters.

Estimation of the number of sexual generations from hybridization in *B.g. triticale*. We estimated the number of sexual generations after hybridization for each *B.g. triticale* isolate using the following formula modified from Stukenbrock *et al.*[8]

$$\text{sexual generations (SG)} = ((\text{NR} \times \text{BGSS})/(2 \times \text{BGSG})) \times 0.22 \quad (1)$$

where NR is the number of recombination breakpoints, BGSS is the average length of *B.g. secalis* segments after two backcrosses (3.3 Mb) and BGSG is the amount of *B.g. secalis* genome in the *B.g. triticale* isolate in base pairs.

The original formulation in Stukenbrock *et al.*[8] was

$$\text{SG} = (\text{ALRWFH})/(\text{ALORW}) \quad (2)$$

where ALRWFH is the average length of recombinant windows after the first hybridization and ALORW is the average length of the observed recombinant window.

We used equation (2) for the calculation of BGSS. We estimated the recombination rate from the genetic consensus map of *B.g. tritici*, which has a length of approximately 1,800 cM (ref. 28). Thus, one expects one recombination event every 10 Mb in a sexual cycle. In turn, after the first hybridization event, we expect segments of *B.g. secalis* genome sequence with an average length of 10 Mb. In the first backcross, these segments recombine once on average, resulting in parental segments of 5 Mb in length. With the second backcross, each window recombines 0.5 times on average. This results in BGSS being on average 3,333,333 bp for *B.g. secalis* segments in *B.g. triticale* after two backcrosses.

In each following generation, *B.g. triticale* isolates mate with each other, and the probability that a new recombination event can be observed corresponds to the probability that recombination occurs in a region that was inherited from *B.g. secalis* in one isolate and *B.g. tritici* in the other. Assuming a proportion of 12.5% for the *B.g. secalis* genome in *B.g. triticale* isolates, this is equal to 0.22 (that is, $0.125 \times 0.875 \times 2$).

Adding these terms, equation (2) becomes

$$\text{SG} = (\text{BGSS}/(\text{average length of } B.g.\ secalis \text{ windows})) \times 0.22 \quad (3)$$

Because the *B.g. tritici* reference genome sequence is fragmented (owing to its extremely high repeat content) (**Supplementary Note**), we cannot use the average length of *B.g. secalis* windows (most of them would be delimited by the end of the contig). However, the detection of recombination breakpoints (NR) is not affected by contig size. The relationship between the average length of *B.g. secalis* windows and NR is given by

$$\text{average length of } B.g.\ secalis \text{ windows} = (2 \times \text{BGSG}/\text{NR}) \quad (4)$$

If we plug equation (4) into equation (3), we obtain equation (1).

We defined putative recombination breakpoints as the borders between regions that contained fixed substitutions inherited from the *B.g. secalis* parent and regions that contained fixed substitutions from the *B.g. tritici* parent. Here we excluded stretches of fixed polymorphisms that are likely the result of gene conversion events. Gene conversions typically affect fragments of less than 1,000 bp in size. Thus, we ignored groups of polymorphisms of a different parental genotype if they were clustered in a region of less than 1,000 bp in size or single substitutions of one genotype in a region otherwise originating from the other genotype.

The proportion of the genome that was contributed by *B.g. secalis* (BGSG) was calculated by adding up the sizes of sequence contigs that only contained polymorphisms of the *B.g. secalis* genotype (**Supplementary Note**).

28. Bourras, S. *et al.* Multiple avirulence loci and allele-specific effector recognition control the Pm3 race-specific resistance of wheat to powdery mildew. *Plant Cell* **27**, 2991–3012 (2015).
29. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
30. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
31. Lipka, A.E. *et al.* GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**, 2397–2399 (2012).
32. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013).
33. Wu, T.D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010).
34. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
35. Hacquard, S. *et al.* Mosaic genome structure of the barley powdery mildew pathogen and conservation of transcriptional programs in divergent hosts. *Proc. Natl. Acad. Sci. USA* **110**, E2219–E2228 (2013).
36. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
37. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
38. Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math Life Sci.* (American Mathematical Society) **17**, 57–86 (1986).
39. Yang, Z. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol. Biol. Evol.* **10**, 1396–1401 (1993).
40. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994).
41. Junier, T. & Zdobnov, E.M. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* **26**, 1669–1670 (2010).
42. Abascal, F., Zardoya, R. & Telford, M.J. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **38**, W7–W13 (2010).
43. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
44. Huelsenbeck, J.P., Larget, B. & Alfaro, M.E. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Mol. Biol. Evol.* **21**, 1123–1133 (2004).
45. Wyand, R.A. & Brown, J.K. Genetic and forma specialis diversity in *Blumeria graminis* of cereals and its implications for host-pathogen co-evolution. *Mol. Plant Pathol.* **4**, 187–198 (2003).
46. Lepage, T., Bryant, D., Philippe, H. & Lartillot, N. A general comparison of relaxed molecular clock models. *Mol. Biol. Evol.* **24**, 2669–2680 (2007).
47. Parlange, F. *et al.* A major invasion of transposable elements accounts for the large size of the *Blumeria graminis* f.sp. *tritici* genome. *Funct. Integr. Genomics* **11**, 671–677 (2011).
48. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
49. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 2014).
50. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
51. Akaike, H. A new look on the statistical model dentification. *IEE Transaction on Automatic Control* **19**, 716–723 (1974).
52. Schwarz, G.E. Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464 (1978).

© 2016 Nature America, Inc. All rights reserved.