the model of molecular evolution that best fitted the data was determined using jModeltest (Posada, 2008). Under the Akaike information criterion (AIC), the data were best fitted by the TPM1uf + G model which assumes equals rates for three types of substitutions (see jModeltest manual for details) with a gamma distribution (G) of site-specific rates. The general time reversible (GTR) model with a gamma distribution (G) of site-specific rates was used in the Bayesian analysis because it is the most similar model to TPM1uf + G available in MrBayes 3.1, and it was ranked 8th among the 88 models compared with the AIC. Priors of state frequencies were left at default settings and Markov chains were initiated from a random tree and run until the average standard deviation of split frequencies remained below 0.01, that is, 600 000 iterations; posterior probabilities were derived from 9000 post-burn-in trees.

To test whether disease status was correlated with life-span (annual vs perennial) while controlling for the plant phylogeny, a continuous Markov model in a maximum likelihood framework was used, as described by Pagel (1994) and implemented in the program Mesquite 2.6 (Maddison & Maddison, 2009) with 1000 simulations. Evidence of phylogenetic signal for the discrete characters of annual vs perennial life-span and presence or absence of disease were tested in Mesquite 2.6 by comparing the observed state transition steps against a simulated distribution of state transition steps in which the character states were shuffled randomly among taxa (Maddison & Slatkin, 1991). In this analysis, a smaller number of observed state transition steps than expected from the randomized distribution (based upon 95% confidence intervals) would indicate that the discrete character is determined significantly by phylogenetic history; the simulated distribution of state transition steps was based on 1000 iterations. These analyses were intended to assess the distribution of plant life-span and disease status rather than to provide systematic revisions to the group, which are underway elsewhere.

Host and pathogen distributions

Because annual species were not found to be diseased (see Results section), we focused on the distributions of perennial species. Information on the geographical distributions of perennial *Silene* species in Europe was obtained from the Atlas Florae Europaeae Database (AFE) (http://www.fmnh.helsinki.fi/english/botany/afe/publishing/database.htm). These data were used to generate species richness maps. For comparison with the perennial species richness map, 10 European *Silene* species were chosen with the most significant binomial distribution probabilities for positive deviations from the expected number of diseased specimens, and 10 European *Silene* species were chosen with the largest numbers of examined specimens where no disease was found. This strategy was used to maximize confidence in assigning species to the two categories of plants with regard to observed rates of disease; the number of species to include in the heavily diseased category was arbitrarily set at a positive binomial distribution probability of 0.05, totaling 10 species in the AFE database, and an equal number of disease-free species was then also included. In addition, 10 annual *Silene* species were chosen with the largest numbers of examined specimens and used to generate a European distribution map. Among the European perennial species the correlation between the size of geographical ranges and the positive or negative deviation from expected numbers of diseased specimens was tested using the Spearman's rank test in spss. The geographical range for each species was determined from the total number of 50 × 50 km grid points occupied by a species in the AFE database. Locality data recorded from herbarium labels of diseased specimens were used to generate a global distribution map of anther-smut disease found in the survey.

Pathogen DNA sampling and analysis

The ability to obtain *Microbotryum* from herbarium samples raises the question of whether it is still possible to isolate DNA and characterize relationships among these samples, especially as some of these specimens would be difficult to resample in the wild. We therefore tested whether DNA could be obtained from herbarium collected smut. We included anther-smut samples infecting the host genus *Calandrinia*, because recent phylogenetic studies (Le Gac *et al.*, 2007) suggested that the anther smut from this genus (in the family Portulacaceae) falls within the clade containing anther smuts from the genus *Silene*. We therefore sought to confirm this in our sampling of anther smut from *Calandrinia* based on a larger number of herbarium specimens.

During the herbarium survey, samples of the infected anthers were collected, with permission, from some diseased specimens (see Fig. 4). DNA was extracted from infected anthers using the DNeasy Mini Kit (Qiagen), and PCR was performed using primers that anneal to the variable regions of the internal transcribed spacer region of the nuclear rRNA genes to amplify only *Microbotryum* DNA (intraITS forward: 5′-CTGTTTAACCAGGGCGTGAC; intraITS reverse: 5′-TGATCTCGAAGGTTAGGATGC). Accession numbers are available in Table S2. Field-collected material of *Microbotryum* from several hosts was also included as positive controls (see Fig. 4). DNA sequences were aligned using ClustalW, and phylogenies were reconstructed with the mega 4.0 software by maximum parsimony analysis using the CNI heuristic search option, 100 random additions of sequences, and 1000 bootstrap pseudoreplicates. Bayesian posterior probabilities for support of tree topology were determined as described in the section on Phylogenetic basis of disease occurrence.