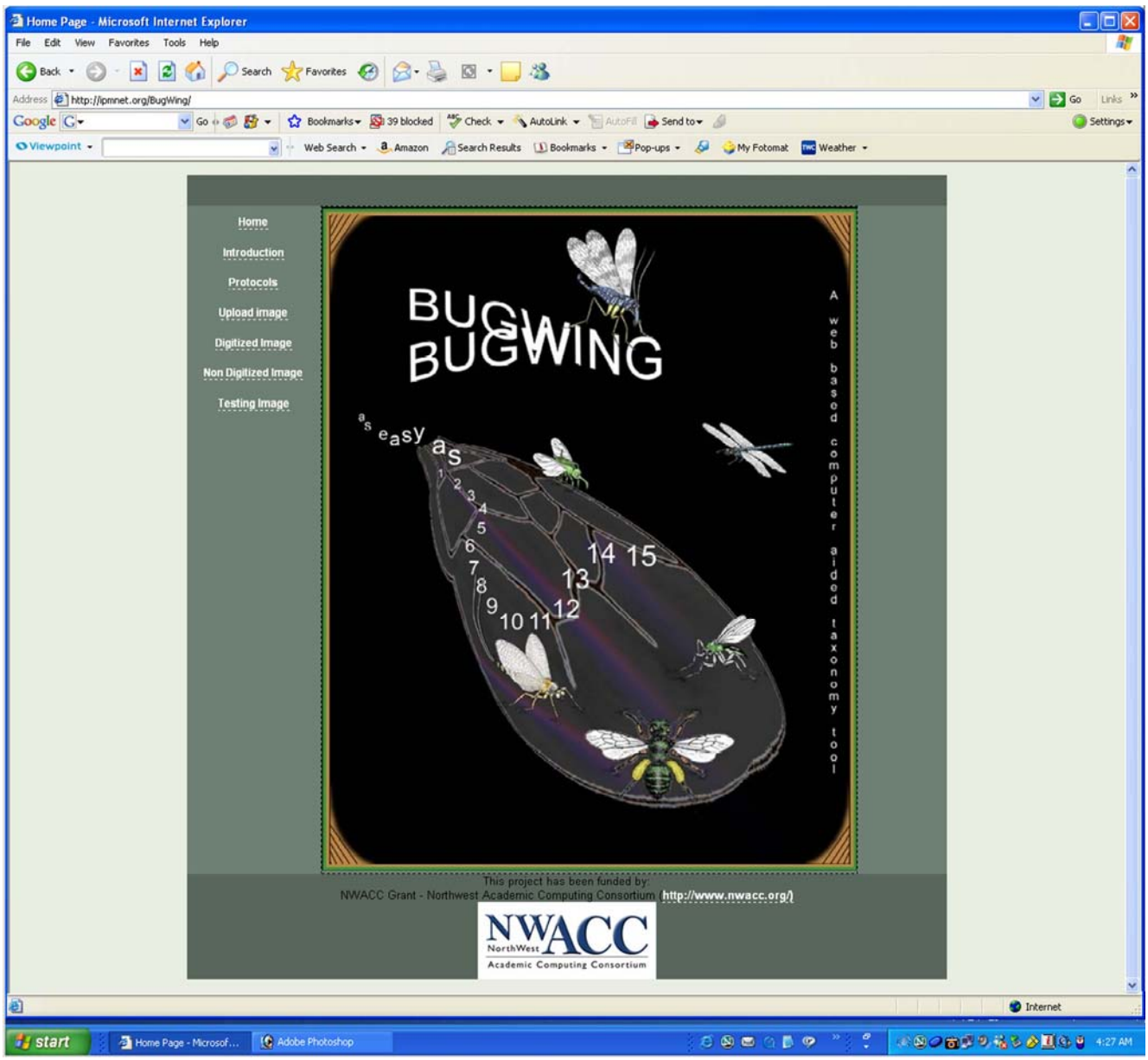
Final Report: NWACC 2007 Proposal April 18, 2008

BugWing@, a web-based computer aided insect identification tool

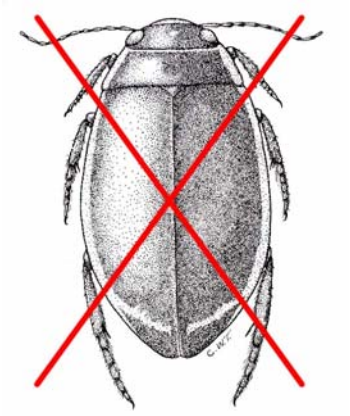
Andrew Moldenke, Dept of Botany and Plant Physiology, Oregon State University, 97331 moldenka@science.oregonstate.edu
Collaborator: Carolyn VerLinden, Dept of Botany and PlantPathology, Oregon State University, Corvallis OR 97331 verlindc@science.oregonstate.edu

Project Website: (http://ipmnet.org/bugwing)
Dedicated Server: "BUGWING" is hosted by **http://ipmnet.org/bugwing** and maintained by Integrated Plant Protection Center at OSU**.**

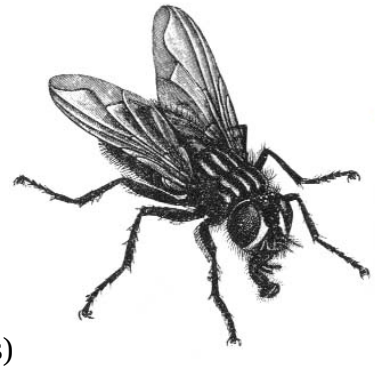


Project Goals:
(Why did we develop the system? What is it designed to do?)

There are so many species of insects that even experts find it difficult and very time-consuming to identify them. An untrained member of the public that has just caught an interesting bug has really no chance at all. However, the adults of most insect species are flighted and the structure of the wing offers many clues to the identity of the insect. The general overall pattern of the number of veins and their intersections usually allows ~~a~~ precise identification to the level of insect order and family. The precise distances between the vein junctions usually allow identification to the genus or species level of resolution. Thus, a semi-automated computer-assisted algorithm that compares patterns of insect wing structures will allow the novice to get a fairly good idea of what type of insect he/she has just collected, and will also allow the professional ecologist to achieve fine scale resolution of ecological samples containing many types of species. While this works for a great number of insects, not all wings lend themselves to this system: very tiny insects with too few wing veins or those with opaque forewings are not candidates for this identification program.



Beetle
(non-transparent wings)



Fly
system designed for transparent, simple pattern

Parasitic wasp
transparent, too few veins

With our identification system all that is necessary for the user is to take a picture of the insect wing (left- or right-hand front wing) with a digital camera, and all that is necessary for us is to provide enough pictures of previously identified specimens in the database for an accurate identification.

Our system was developed for 2 types of users:

1) Middle- and high-school biology students participating in our Ecoplexity hands-on training program (http://ecoplexity.org). In this program students in

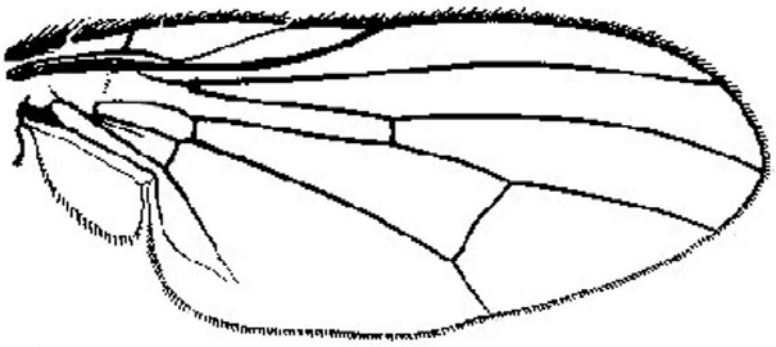
OR, WA, CA, CO, NM, AZ and Puerto Rico conduct ecological monitoring studies under the direction of Long Term Ecological Research (NSF-LTER) scientists. The protocols they use are the same as those used by scientists at all LTER sites and the data are then comparable with other student teams across the country as well as their local ecologists. Identifications during the school terms need to be as precise as possible and digital images and specimens need to be archived for future critical analysis. Correct identifications allow students to compare results during subsequent years, and precise (or as precise as possible) identifications permit the collaborating scientists to verify and enter the data as quickly as possible. Precision is a function of the extent of the databased image collection, but students are already using the system at this proof-of-concept stage of development.

2.) Ecosystem ecologists at the LTER sites in OR CO, AZ, NM and PR (& soonto be International LTER sites in East Asia). Most ecological protocols involving monitoring and collection yield an enormous diversity of insects. This creates a SERIOUS procedural problem in analysis which is generally solved in one of two ways: 1) the number and size of samples is greatly reduced which seriously impacts statistical analysis and predictability; and/or 2) the taxonomic resolution of the sorting process is greatly decreased, which in many instances limits the resultant descriptive power. Any system that could permit speedy and robust identifications of a wide variety of insect types would revolutionize both ecological monitoring and synecological research. Though BugWing© is only at proof-of-concept stage of development, we are working on defining the resolution possible with different types of insect groups (for instance how well it can distinguish between very closely related species within a single genus of bees) and subsequent linkage to BugBytesNorthwest©, with its databanks of expanded information for each species, including ecological roles, literature, etc. (for instance how can we build in a seamless transition from BugWing's identification keys to BugBytesNW's ability to locate pertinent scientific literature, etc. on any identified species.)

How the system works:

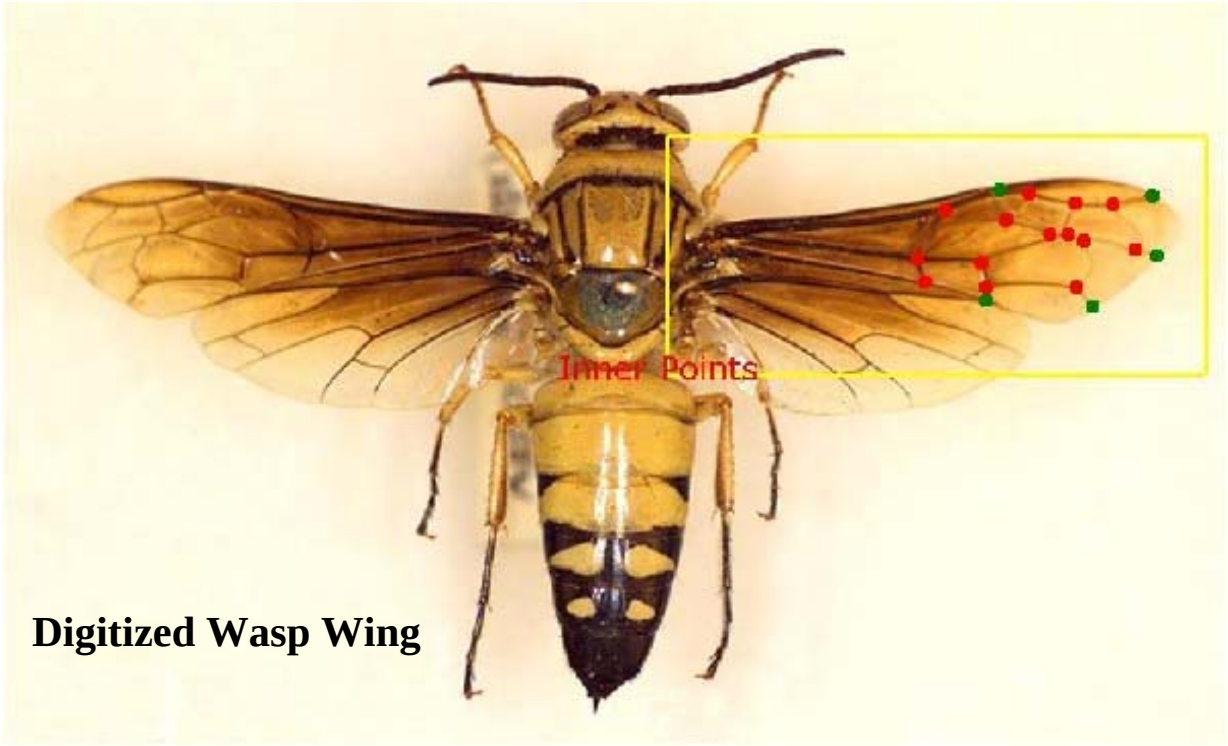
With current technological advances in digital camera optics, it is nearly universally possible to take a digital image of an insect wing without fancy microscopes or macro-lenses. 'Everyday' digital cameras have both the focal range and resolution capability to capture an insect wing image and enlarge it sufficiently to be examined by the BugWing© system. Once the image is captured, it has only to be saved as a .jpg file.

Undigitized wasp wing

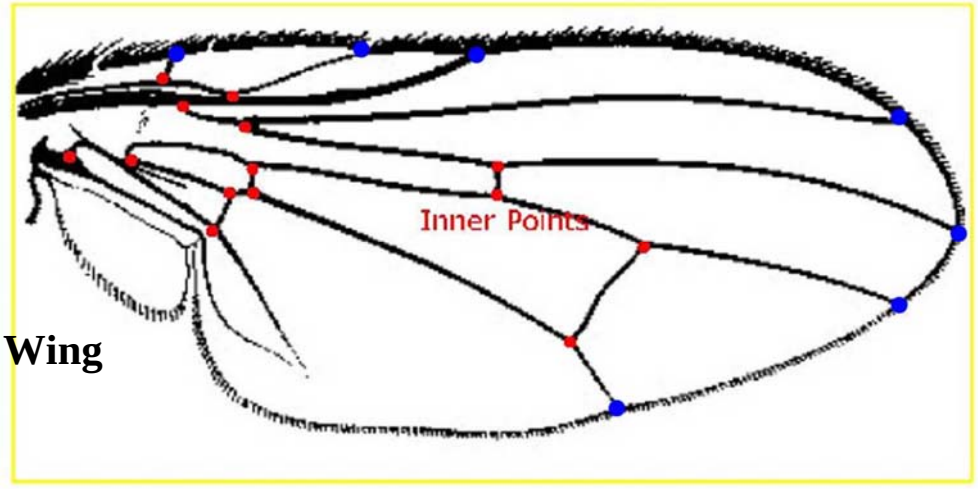


Undigitized fly wing

The on-line availability of BugWing© allows any end-user to process an image at his/her convenience. (1) The image is uploaded into the BugWing© program; (2) instructions for processing are followed (see “Protocols”); (3) the image is digitized and saved;

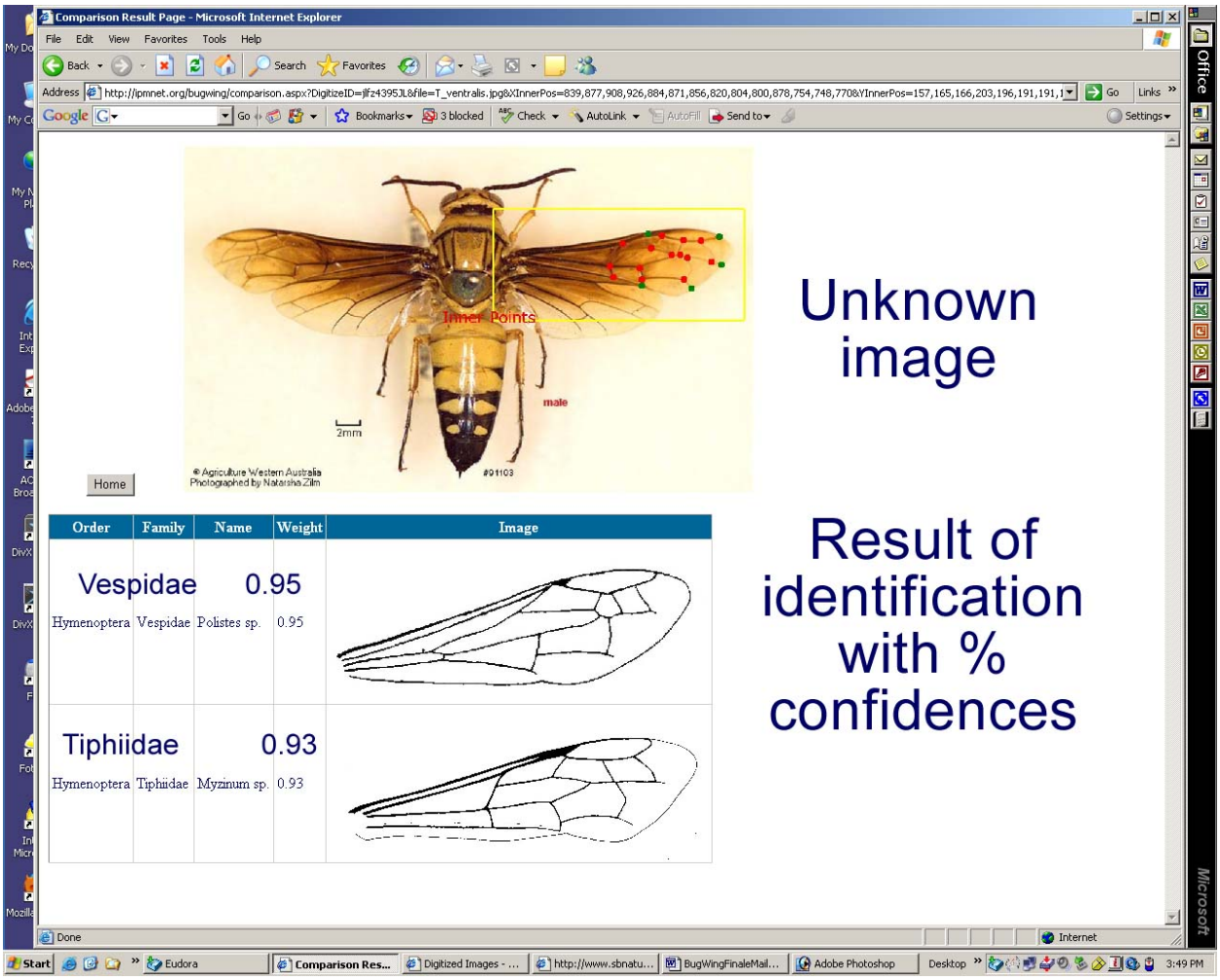


Digitized Wasp Wing



Digitized Fly Wing

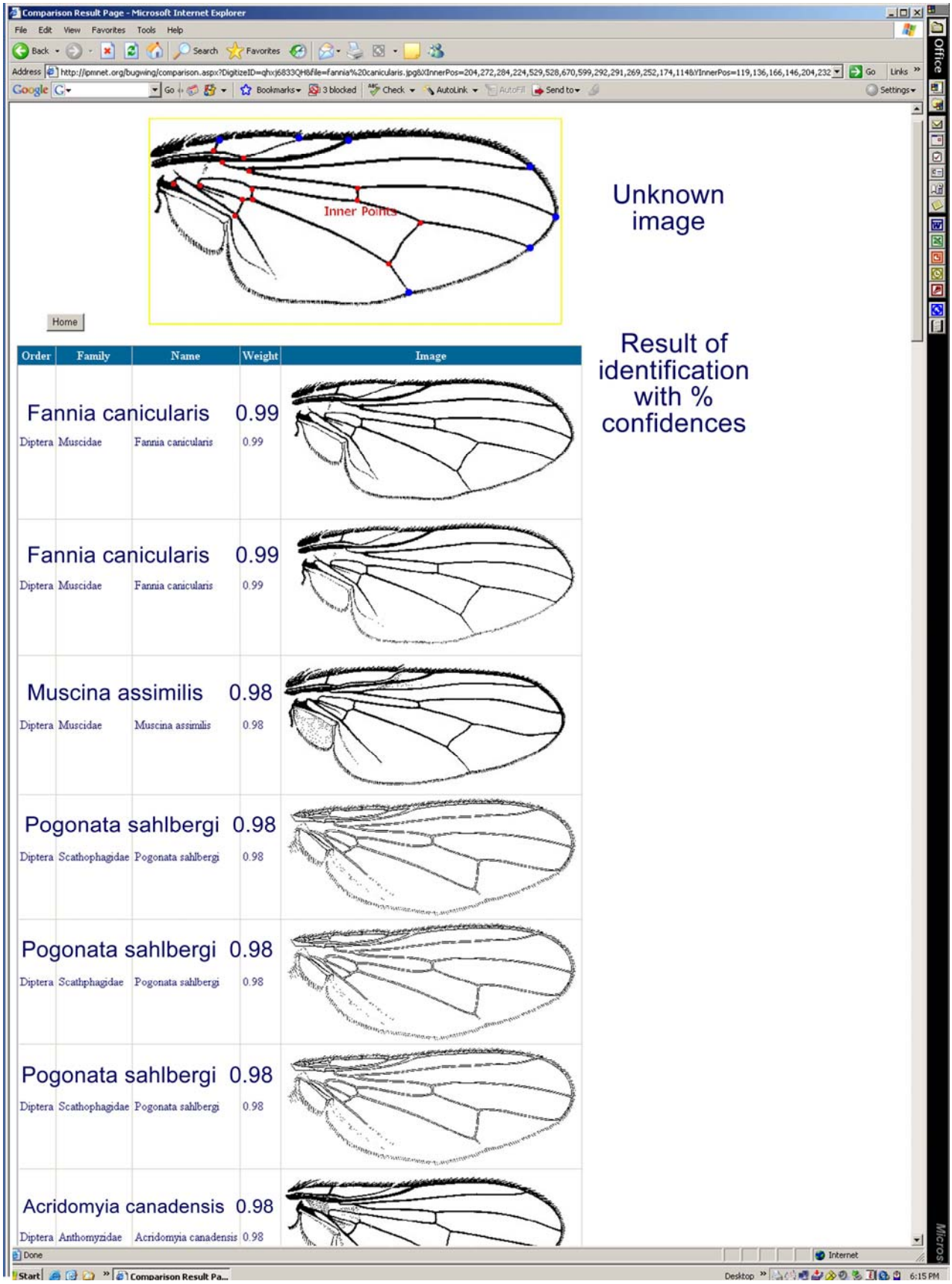
and the (4) “Identify” button clicked. The user is then presented with a ranked series of 10 images from the database which most closely resemble the digitized image just submitted. For any result from 100% confidence to something less, the calculation of similarity is based solely on the amount of information in the database of images. This database will always be somewhat incomplete, so we are working on providing the user with the information for interpreting any series of partially matched images.



In order to achieve the highest degree of database precision, we have entered many hundreds of line drawings of insect wings from classic entomological research papers/books. This allows us to approximate the entire range of variability of possible wing structures potentially encountered on the scale of the North American continent. Subsequent analysis of each wing type allows us to determine whether it represents a structure diagnostic at the family, subfamily, tribe, genus or species level; in general, these representative wings are characteristic of several closely related genera.

We are also extremely interested in the ultimate level of resolution that BugWing© can provide. To this end we are testing closely related species within 2 genera of bees, *Bombus* and *Melissodes*, both of which are characterized by being composed of many different species. Even very closely related species can be told apart by BugWing but our

current “one algorithm fits all” protocol may not ultimately turn out to be the most efficient design to distinguish between all the different types of wings.



The goal of this project has been to develop an online automatic identification of insects. Because this has been a proof of concept project, the morphometric approach has focused only on the insect wing (ignoring all other types of morphologic characters). Several interrelated morphological characteristics are derived by measuring the network pattern, such as distance between vein junctions and area of wing skin enclosed in a wing cell. These measurements represent a unique feature of each individual, just like human's fingerprint.

In this project, we have designed a typical 3-tier application. The three major parts of application are the presentation interface, the server-side programs, and the knowledge database. We have used the World Wide Web as the user interface. It contains application-specific entry forms for users to upload images, interact with the application, and gain knowledge from the database. The server-side programs respond for user requests from the web pages. In turn, they convert the digital image into morphometric data, determine what data are needed (and where they are located) and conduct multivariate analysis to classify the specimen into a taxonomical group. The third tier includes an iterative knowledge base system to manage insect wing data. There are two types of database, contextual information and testing data. The former represents encyclopedic and textual data which are built based upon the insect data of the Arthropod Collection at Oregon State University . The database provides a lot of information on specimens (landmark coordinates, number of junctions, vein lengths, etc.) that can be useful for classification and identification. The latter stores morphological data derived from specimen images that users uploaded to the server.

Extent to which goals have been met:
In theory, there are several distinct components needed to develop this system:

1) (Computational) Writing the algorithm to accomplish the basic identification process. **Accomplished**.
2) (Computational)_Writing the algorithm to accomplish the image manipulation (uploading, digitizing). **Accomplished.**
3) (Computational) Establishing web accessibility and long-term server commitments. **Accomplished**
4) (Biological) Experimenting with a number of different protocols to prioritize the morphometric measurements. **Accomplished** – simplicity and clarity required using only a single basic protocol for a diversity of structures.
5) (Biological) Determining the necessity for separate protocols for different wing types. General type – **Accomplished**; modifications for bees, dragonflies and damselflies **– Accomplished**; modifications for Neuroptera **not yet attempted**.
6) (Biological) Testing the ability to perform the initial stage of identification to basic insect family-level. **In process**, mostly successful.
7) (Biological) Testing the ability of the system to distinguish between closely related species**. In process**, images assembled, testing still underway – long-term iterative process separately required for many different groups.
8) (Computational) Testing the variability inherent as different end-users digitize the same images. **In process**, both using ‘simplified’ line drawings of wings and images of actual specimens taken with standard digital cameras.

9) (Databasing) Textbook images of all different major insect wing types. Entering in database – **Accomplished**; digitizing images – **In process**.
10) (Databasing) Acquisition of images from OSU Arthropod Collection – **Pending**, digital camera bought with grant funds, collaboration with student volunteers to start during summer quarter. First insect group to be imaged will be bees.
11) (Computational) Increased user-friendliness - **Ongoing**

Impact and future plans:

This project offers a number of possibilities to apply information technology over a broad range of applications benefiting distance learning. The web interface provides not only a way for communication between students and instructors, but also easy access for remote studies. The interaction/connection between instructors and students will not have to be as frequent, so students can conduct self-learning or independent learning on insect identification from any web browser.

The traditional tools (both hardware and software) for morphometrical analyses are expensive, including microscopes, digital cameras, image processing software, and a statistical package. Most institutes can afford only a limited number of tools, and ask their students to operate the tools in a laboratory. Because of the popularity and accessibility of the World Wide Web, this application can significantly reduce the financial burden for the course instructors and students. It will also be valuable for entomologists who are interested in morphological identification and run the tools via a web browser for free.

Even though the database represents only a fraction of the insects in North America, we will start to test its utility with students in the Advanced Science Program at **Lincoln Savage Middle School** in Grants Pass, OR, during spring semester. Students are already doing environmental sampling and are taking digital pictures of the specimens collected. We will use the results of processing their specimens as an independent test of project utility. During the summer it will be tested again by teachers from 5 states in the **Exoplexity Program**, and then further tested by their students during school semesters.

Collaboration and related Website:

Our related project, **BugBytes** (http://ipmnet.org/ent3/bugbytes/), will be incorporated into the procedure to reflect the synergy of a new inter-site ecology education program, Teaching Ecological Complexity, and funded by National Science Foundation. Our collaborators include Dr. Marion Dresner (Portland State University), Stephanie Bestelmeyer (New Mexico State University), John Moore (Colorado State University), Jorge Ortiz (University of Puerto Rico) and Monica Elser (Arizona State University). BugBytes is a web-based insect identification system based on the matrix of characters rather than the typically used dichotomous key. Thousands of insect characters and taxonomy are stored in a relational database. The identification web page is created based on users' interaction to the web page. Although BugBytes contains a great amount of insect data, it doesn't have the function to automatically convert morphological characters to quantitative data. BUGWING will be incorporated into the BugBytes project and will serve as a morphometric tool for the users to measure morphological data from specimens generated in public high- and middle-school science projects supported by the collaborating universities.

Dissemination and Publicity:

As we populate the database with more images and are able to interpret more fully exactly which species as yet unentered into the database are represented by the databased taxa, the system will become more widely used by the **school children** in our outreach programs of the LTER system. We hope to attain a level of resolution that will identify all insects to both family-level resolution and specify what functional role it plays in the ecosystem. The goal of our educational program is to demonstrate both high levels of insect diversity AND to integrate them into an understanding of foodwebs and ecological complexity.

As for the professional insect ecologist we hope that by exploring the extent to which closely-related species can be distinguished (e.g., bees) that we will be able to demonstrate it publicly at the next meeting of the **International Congress of Entomology** in Durban, RSA, July 2008.

Impact:

There are a vast number of teaching websites, most of which use a content-based web design; i.e., web pages are the media used to present data. Very few teaching web designs treat a website as a computer application, so that students can learn by interacting with the application. The differences between web pages and web applications are subtle and depend on the level of a user's input to affect the outputs from the website. Most web designs are so-called static web pages, created with fixed HTML text and coding. Any change to the page requires editing the page from the instructor on the server side. Most instructors prefer to design teaching web sites with the static web pages because the designing concept is simple, similar to creating a PowerPoint file for displaying slides. In contrast, web applications only contain rich content about ecological models but also possess functions that will allow students to using the tools and gain hands-on experience. A web application, like any software product, is a task-oriented web site that performs some useful functions. In our web application, students can use this web application just like they conduct morphological identification in the laboratory.

Closing Financial Statement:

The entire budget has been spent in accordance with NWACC policy. Final budget accounting will be officially submitted by the University Contracts Officer at a later date.

Matching funds and computer/networking support will be provided by Integrated Plant Protection Center via OSU. We are currently supported by a grant from the National Council for Air and Stream Improvement who are interested in developing more inclusive assays for stream health (current analyses utilize only mayflies, stoneflies and caddisflies – not true flies because of the procedural difficulty of identifying them; true flies usually encompass far more than half the insects in a stream sample). We will also be approaching EPA in the near future for grant funds as well for image analysis development (in collaboration with Tom Dietterich, OSU).