Deep infrared studies of massive high redshift galaxies

Labbé, I.

CHAPTER

SIX

IRAC Mid-Infrared Imaging of Red Galaxies at $z > 2$ new constraints on age, dust, and mass

ABSTRACT

We present deep $3.6 - 8$ micron imaging with IRAC on the Spitzer Space Telescope of a population of galaxies with red rest-frame optical colors at $z > 2$. The 13 distant red galaxies (DRGs) were selected in the field of the Hubble Deep Field South on the simple color criterion $J_s - K_s > 2.3$ and we compare their properties to those of 23 Lyman Break Galaxies (LBGs or U-dropouts) at $z \sim 2.5$ in the same field. The new IRAC data reaches rest-frame NIR wavelengths, which are crucial in determining the nature of these galaxies. We are able to uniquely identify 3 out of 11 DRGs as old passively evolving systems at $z \sim 2.5$. The others are heavily reddened star-forming galaxies, for which we are now better able to distinguish between the effects of age and dust. Furthermore, the rest-frame NIR data allow more robust estimates of the stellar mass and stellar mass-to-light ratios (M/L_K). We find that in a mass-selected sample DRGs contribute $1.5-2\times$ as much as the LBGs to the cosmic stellar mass density at $2 < z < 3.5$. Also, at a given rest-frame K luminosity the red galaxies are twice as massive with average stellar masses $\sim 10^{11}\ M_\odot$, and their M/L_K mass-to-light ratios exhibit only 1/3 of the scatter compared to the U-dropouts. This is consistent with a picture where DRGs are more massive, more evolved, and have started forming at higher redshift than most LBGs. We find no evidence for a substantial AGN contribution to the observed optical-MIR SEDs.

Ivo Labbé, Jiasheng Huang, Marijn Franx, Gregory Rudnick, P. Barmby, Emanuele Daddi, Pieter G. van Dokkum, Giovanni G. Fazio. Natascha M. Förster Schreiber, Konrad Kuijken, Alan F. Moorwood, Hans-Walter Rix, Huub Röttgering, Ignacio Trujillo, Arjen van der Wel, Paul van der Werf, & Lottie van Starkenburg

1 Introduction

MAPPING the properties of massive galaxies as a function of redshift provides strong tests for theories of galaxy formation, as their build-up can be directly probed from high redshift to the present epoch. But while the theories describing the growth of large-scale dark-matter structure are thought to be well-constrained (Freedman et al. 2001; Efstathiou et al. 2002; Spergel et al. 2003), the formation history of the stars inside the dark-matter halos is still poorly understood. Direct observations are essential for progress on this front.

Until recently, it was difficult observe statistically meaningful samples of massive high-redshift galaxies. The best-studied samples are selected on the rest-frame UV light through the Lyman Break technique (LBGs; Steidel et al. 1996a,b, 2003; Madau et al. 1996; Giavalisco & Dickinson 2001), which yielded large numbers of relatively low mass, unobscured, star-forming galaxies at $z > 2$ (Papovich, Dickinson, & Ferguson 2001; Shapley et al. 2001). Recent advances in near-infrared (NIR) capabilities on large telescopes are now making it possible to access the rest-frame optical for large numbers of galaxies to $z \sim 3$. The rest-frame optical is already much less sensitive to dust obscuration and on-going star formation than the rest-frame UV, and is expected to be a better (yet imperfect) tracer of stellar mass.

In this context, we started the Faint Infrared Extragalactic Survey (FIRES; Franx et al. 2000): a deep optical-to-infrared multicolor survey of NIR-selected galaxies in two fields. In the deepest field, the Hubble Deep Field South (HDFS), we spent 102 hours of VLT/ISAAC imaging on one pointing in the J_s, H, and K_s-bands resulting in the deepest ground-based NIR imaging, and the deepest $K-$band imaging to date, even from space (Labbé et al. 2003). We selected high-redshift galaxies with the simple color criterion $J_s - K_s > 2.3$, designed to isolate galaxies at $2 < z < 4$ with a prominent Balmer- or 4000 Å break (see Franx et al. 2003). We find these Distant Red Galaxies (DRGs) at high surface densities ~ 3 arcmin^{-1} to $K = 22.5$, with space densities about half of that of LBGs selected from ground-based imaging down to $R = 25.5$.

Previous studies using broadband optical-to-NIR SED fitting and optical/NIR spectroscopy have suggested that DRGs, at a given rest-frame optical luminosity, have higher ages, contain more dust, and are more massive than LBGs (Franx et al. 2003; van Dokkum et al. 2004; Förster Schreiber et al. 2004), and they may contribute comparably to the cosmic stellar mass density (Franx et al. 2003; Rudnick et al. 2003). However, the nature of their red colors is still poorly understood, and the masses are somewhat uncertain, raising many questions. Are they all truly old, or are some also very young and very dusty? What is the fraction of passively, evolving "dead" systems? How much do the DRGs contribute to the stellar mass density in a mass-selected sample? And how do they relate to the blue Lyman break galaxies. Finally, what is their role in the formation and evolution of massive galaxies?

To shed light on all of these questions, we need imaging in the rest-frame NIR, including the rest-frame $K-$band. Here, we present deep MIR imaging with IRAC on the Spitzer Space Telescope of a sample of distant red galaxies in the field of the HDFS. In combination with the wealth of existing deep imaging from HST/WFPC2 and VLT/ISAAC, the IRAC data promises to establish more robust stellar masses and mass-to-light ratios, and may help to reduce the degeneracies between age and dust in modeling of the broadband spectral energy distributions (SEDs) (see, e.g., Papovich, Dickinson, & Ferguson 2001; Shapley et al. 2001). Where necessary, we assume $\Omega_M = 0.3, \Omega_\Lambda = 0.7$, and $H_0 = 70$ km s^{-1}Mpc^{-1}. The magnitudes are given on the Vega system.

2 The Observations, Photometry, and Sample Selection

We have observed the HDFS WFPC2 field with the IRAC camera (Fazio et al. 2004) on the Spitzer Space Telescope, integrating 1 hour each in the 3.6, 4.5, 5.8, and 8 micron bands. The full observations and data reduction are described in Labbé et al. (2004), but we give an abbreviated outline here.

The IRAC images were taken in the 4 broadband MIR filters at 3.6, 4.5, 5.8, and 8 microns over a $5'$ x $5'$ field of view. The pixels are $\approx 1.2''$ in size. The observations in the field of the HDFS were taken on May 26 2004 (GTO program 214) and were split into dithered frames of 200s each, except for the 8μ band where the frame time was 50s.

We used the Basic Calibrated Data (BCD) as provided by the Spitzer Science Center pipeline, and refined the astrometry of the individual frames with 2MASS sources. We rejected cosmic rays, corrected for known artifacts such as column pulldown, and muxbleed, and accounted for the "first frame effect" by subtracting a median stacked image from the individual frames. Finally, we corrected the frames for geometric distortion, projected them on the existing ISAAC $K-$band image (Labbé et al. 2003), and average-combined them. We compared the result to a median combined image to make sure that we excluded all cosmic rays.

The pixel scale of the final images is $0\farcs36$ per pixel, about 1/3 of the original IRAC scale and equal to $3\times$ the scale of the ISAAC image. The limiting depths at 3.6, 4.5, 5.8, and 8 micron are 22.2, 21.3, 18.9, and 18.2 mag, respectively (derived from the 5σ effective flux dispersion in 3 arcsec diameter apertures). The image quality ranges from 2 - 2.4 arcseconds FWHM and is best in the 4.5μ band. The positional accuracy with respect to sources in the ISAAC $K-$band is better than $0\farcs2$ across the field. From hereon, we only consider the deepest central part of the MIR imaging overlapping the $2.5' \times 2.5'$ ISAAC field.

To achieve consistent photometry across the MIR IRAC bands and the existing NIR catalog[1], we carefully matched the point-spread-function (PSF) of the $K_s, 3.6, 4.5$, and 5.8 band to the 8μ band, where the image quality was worst. We

[1]NIR images and catalogs are available from http://www.strw.leidenuniv.nl/~fires

deconvolved a selection of bright stars in the 8 micron image with a selection of bright stars in the other maps, providing us with the required PSF-match kernels. We convolved the maps to the 8 micron PSF, and checked the quality of the match by dividing stellar growth curves, normalized to 1 at the IRAC zero-point radius of $R = 12.2$ arcseconds. The agreement is better than 3% at radii $R > 2.2$ arcseconds; hence we used a diameter of $D = 4\farcs4$ as our fixed aperture for photometry.

To reduce the effects of confusion, we used the deep ISAAC $K-$band image to model and subtract neighbouring sources. The image quality in the $K-$band is $\sim 0\farcs5$, so confusion is less of an issue. We proceeded by convolving each K-band source in isolation to the IRAC PSF, and fitted them simultaneously to each individual IRAC map, leaving only the fluxes as free parameters. We did not use the model fluxes directly; instead, we use the models to subtract the confusing neigbours next to the sources of interest. We then proceeded with normal aperture photometry in a $D = 4\farcs4$ diameter. We combined the IRAC fluxes with existing optical/NIR photometry in $D = 2''$ apertures from the catalog published by (Labbé et al. 2003). Thus to obtain consistent colors, we applied an aperture correction to the IRAC fluxes which was the ratio of the original $K-$band flux in the $D = 2''$ aperture and the $K-$band flux in the PSF-matched $D = 4\farcs4$ aperture. Finally, we assumed a minimum photometric error of 10% reflecting various calibration uncertainties. The end result is a fairly homogenous photometric catalog spanning the observed optical-to-MIR in 11 filters.

The primary sample of interest is comprised of the DRGs in the field of the HDFS which were selected on $J_s - K_s > 2.3$ and $K < 22.5$ (see Franx et al. 2003). The sample comprises 14 galaxies in the redshift range $1.9 < z < 3.8$, which were all detected at 3.6, and 4.5 micron. Two galaxies have only upper limits at 5.8 and 8.0 micron. One of these was excluded from further analysis as the deblending procedure was unsuccessful; the source was also confused in the $K-$band. The final sample contains 13 DRGs.

As a comparison sample, we selected Lyman Break Galaxies in the same field from the WFPC2 imaging (Casertano et al. 2000) to the same limit in K, using the criteria of Madau et al. (1996). We note that the photometric system of the WFPC2 is somewhat different from the one adopted for the selection of ground-based LBGs (Steidel et al. 2003). The blue $F300W$ bandpass causes Lyman break galaxies to enter the selection window already at redshifts $z \gtrsim 1.8$ (Giavalisco & Dickinson 2001). As a result the redshift distribution of space-based "U-dropouts" is a better match to that of the DRGs than the groundbased LBGs, or the more recent "BM/BX" objects at $z \sim 2$ (Steidel et al. 2004).

We kept the 37 LBGs that fell in the same redshift range as the DRGs ($1.9 < z < 3.8$). Because of their higher surface density, source confusion played a more important role. Conservatively, we selected only the most isolated, and hence least affected galaxies out of the 37, leaving a final sample of 23. We checked that the median K-band magnitude and the median $I - K$ color of this subsample was

the same as that of the full sample; hence it should be representative. Whenever calculating global averages, we simply scale the average properties of either sample to the original numbers (14 and 37).

To conclude, the median $K-$band magnitude of the DRGs is 21.4 and that of the space-based LBGs is 21.5. The redshift distributions are comparable, and both have a median $z_{phot} = 2.5$. Stars were excluded from the LBG sample using a method detailed in Labbé et al. (2003) and Rudnick et al. (2003). The galaxy identifications througout this paper refer to the published catalog of Labbé et al. (2003), which includes photometric redshifts for all sources.

We determined the rest-frame luminosities and colors by direct interpolation in AB magnitudes between the observed filters. The rest-frame luminosities are sensitive to the uncertainties in the photometric redshifts. We used photometric redshifts based on an algorithm developed by Rudnick et al. (2001, 2003) and spectroscopic redshifts where available. The photometric redshifts are in good agreement with the spectroscopic redshifts, with an average $|z_{phot} - z_{spec}|/(1 + z_{spec}) = 0.05$ for sources at $z > 2$. We checked that the publicly available HYPERZ method (Bolzonella, Miralles, & Pello 2000) gave consistent answers, and we checked that adding the MIR did not change the photometric redshifts significantly. We decided not to use these new photometric redshifts for now, as the differences were small.

3 Mid-Infrared Properties of Red Galaxies at $z > 2$

One of the primary motivations for obtaining the $3.6 - 8$ micron IRAC imaging is to extend the spectral energy distribution (SED) of the DRGs into the rest-frame near-infrared. The rest-frame NIR is essential to understand why these galaxies have colors that are much redder than Lyman Break Galaxies. Is it because they are much more obscured, or because their stellar populations are more evolved? The answer to this question is a crucial step in understanding their mutual relation. Some insight has already been obtained from modeling of the observed optical and near-NIR SED (Förster Schreiber et al. 2004), but it has not been possible to uniquely distinguish old, passively evolving galaxies from very dusty, actively star forming systems.

The rest-frame NIR is expected to seperate very young and extremely reddened ($A_V > 3$) galaxies from older galaxies with much less extinction ($A_V < 1.5$), as in the first case the spectrum is red througout the rest-frame UV and optical, and peaks somewhere around the rest-frame J-band, while in the other case the spectrum peaks in the rest-frame optical.

As a first step we can see if the nature of these objects can be elucidated by a specific combination of two colors alone. We inspect the observed $I - K$ versus $K - 4.5\mu$ color-color diagram of the galaxies (see Figure 1). At a redshift of z=2.5, these colors correspond to the rest-frame $2200 - V$ versus $V - J$ colors. We examine the three DRGs whose optical-to-NIR SEDs were fit significantly better

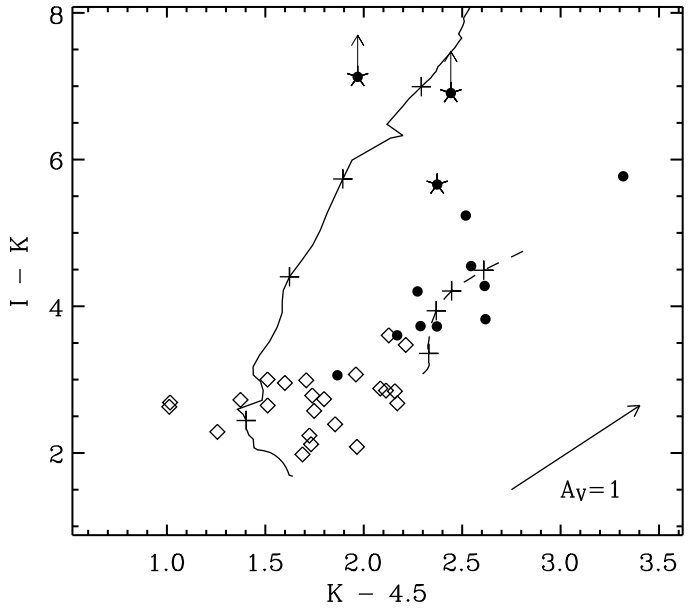

Figure 1 — $I - K$ versus $K - 4.5$ color-color diagram of two samples of $z > 2$ galaxies. Diamonds show Lyman Break galaxies. Filled circles show galaxies selected on the color-criterion $J - K > 2.3$ or Distant Red Galaxies (DRGs). Red galaxies whose optical-to-NIR SED were best fit with unreddened single age bursts are marked with a star. Also shown are time-evolution tracks of Bruzual & Charlot (2003) stellar population models, redshifted to $z = 2.5$. (*solid line*) A single age burst, (*dashed line*) a constant star forming model with a reddening of $A_V = 1.5$. The reddening vector assumes a Calzetti et al. (2000) extinction law.

with unreddened single age bursts than with dusty constant star formation models (indicated by stars in Figure 1); these are amongst the reddest in $I - K$ color. For these three galaxies the best fit single age models predict fairly blue $K - 4.5\mu$ colors. On the other hand, the worse fitting dusty constant star formation models for these galaxies predict them to be the reddest in $K - 4.5\mu$, which they are not. Had we only $K - 4.5\mu$ color, and hence lacked the apriori knowledge of the best-fit SFH, this result implies that we would still have been able to crudely distinguish old passively evolving galaxies from those that are heavily reddened and vigorously forming stars.

To look in more detail at the three candidate old "dead" galaxies we show in Figure 2 their broadband SEDs. Remarkably, single burst model predictions based on the optical and NIR data alone, are directly confirmed by the IRAC $3.6 - 8$ micron observations. For these galaxies we can definitively rule out reddened

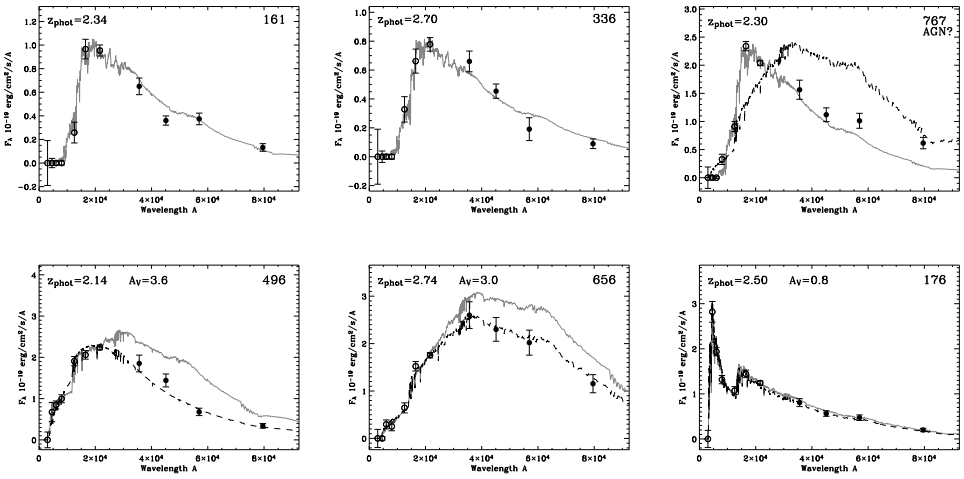

Figure 2 — Comparisons between the predicted IRAC fluxes from fits to the optical-to-NIR SED and the actual IRAC $3.6 - 8$ micron observations. The predictions (*gray solid lines*) are based on Bruzual & Charlot (2003) stellar populations fitted to the optical-to-NIR SEDs (*open circles*). The top row shows the 3 DRGs which, excluding the MIR data, were best-fit with an unreddened single age burst. The observed MIR IRAC fluxes (*filled circles*) provide direct confirmation on their nature. Galaxy 767 shows a flux excess at 8 micron, which cannot be produced by constant star formation and reddening (*black dashed lines*); it is possibly related to AGN activity. The bottom row shows the MIR predictions for galaxies using constant star forming models with Calzetti et al. (2000) dust reddening. For heavily reddened galaxies, the MIR observations were often different from the optical-to-NIR based predictions, reflecting the degeneracy between age and dust in the models. For galaxies with lower dust content $A_V \lesssim 1$, IRAC confirms the predictions. Also drawn is the best-fit to the full SED (*black dashed lines*).

models with constant star formation. Before the MIR imaging, we could only marginally exclude these models from the formal χ^2 of the fit (Franx et al. 2003). We also tried less extreme models with declining star formation rates ($\tau = 300 - 500$ Myr) and dust reddening (Calzetti et al. 2000); we find that such models fit significantly worse than the single age bursts.

One of the galaxies clearly has excess flux at 8 micron. Complex stellar populations with, for example, heavily reddened star formation cannot account for such excess, because they also generate $5.8\mu - 8\mu$ colors that are too blue. It may be related to an obscured active galactic nucleus (AGN).

Also in Figure 2 we show the MIR predictions for galaxies whose optical-to-NIR fluxes were better fit with constant star forming models and reddening. In general the observed MIR flux points do not confirm the predictions, reflecting the well-known degeneracy between age and dust in the models which allows a wide variety of predicted values(see, e.g., Papovich, Dickinson, & Ferguson 2001). This effect is particularly severe for heavily reddened galaxies that are red throughout

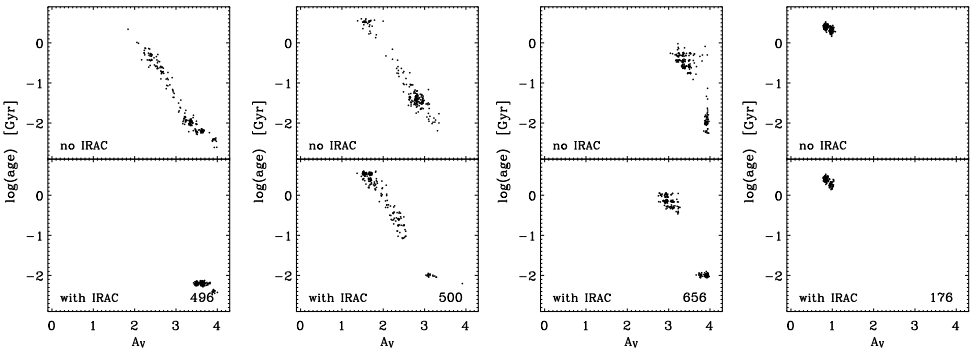

Figure 3 — The range in best-fit extinction and age values of various DRGs, for constant star forming models (Bruzual & Charlot 2003) including dust Calzetti et al (2000). We show separately the solutions obtained excluding, and including the mid-infrared data. The range in values is derived from Monte-Carlo simulations, where the photometry is randomized within the photometric errors, and the fitting procedure is repeated. The improvement of the constraints depends on the nature of the galaxies.

the optical and NIR, but much less so for moderately obscured galaxies that peak at observed $\lambda < 2.2\mu$, e.g. object 176.

The immediate question is how much the MIR fluxes improve the constraints on the models? To investigate this used the HYPERZ package (Bolzonella, Miralles, & Pello 2000), updated with the latest Bruzual & Charlot (2003) templates, to fit stellar population models to the full SED. We used solar metallicity models with a Salpeter IMF ranging from $0.1 - 100\ M_{\odot}$, and we adopted a Calzetti et al. (2000) extinction law. In the fitting we kept the redshifts fixed to the values we derived earlier with an different method (see §2), and we restricted the ages to the age of the Universe at each redshift.

Next, we derived confidence limits on the best-fit values for age and dust through Monte-Carlo simulations. For each galaxy, we generated 200 simulated SEDs by randomizing the photometry within the photometric errors, and we fitted them in the same way as the observed SEDs. We did the simulation twice: one time excluding the MIR fluxes, and the other time including the MIR fluxes. We derived the 1σ limits on each from the central 68% of the distribution. We crudely accounted for systematic errors by multiplying the limits with the square root of the reduced χ^2 of the best-fitting model. We assume that the three free parameters in the model are age, dust, and star formation rate. We show in Figure 3 the simulation output for 4 DRGs spanning a wide range in properties. Three of them were also present in Figure 2.

We can immediately learn several things from the impact of the new MIR data. Firstly, it has become possible to distinguish extremely reddened, very young models ($A_V > 3$) from much older less reddened models ($A_V < 2$) (see galaxy 496 and 500 in Figure 3). Galaxy 496 is now uniquely fit with a very

young, heavily reddened model, implying instant star formation rates of more than $2000 M_{\odot} yr^{-1}$, in the range of the sub-mm selected "SCUBA" sources (e.g., Ivison et al. 2000; Smail et al. 2000). Galaxy 656 is best fit with a very old, very dusty model in both cases. Here the improvement is modest, although best-fit solution has shifted somewhat. It is clear though that even with deep IRAC imaging the age-dust generacy is not completely resolved for all sources. We note that the uncertainties on the IRAC fluxes are mostly systematic, hence better calibration can improve the situation. But ultimately other observations, such as NIR spectroscopy, are needed to place independent constraints. Finally, as already suggested by Figure 2, contraints on models that are dominated by the blue stellar continuum with moderate extinction levels $A_V < 1$ do not improve particularly from the MIR data: the parameters were already well-constrained from the optical and NIR data alone.

In summary, we find for the whole sample of 13 DRGs that 3 galaxies are uniquely identified as old and passively evolving, 7 galaxies are well-fit by dust-reddened and star forming models with a range in extinctions and ages, 1 galaxy can be fit with either model, and 2 galaxies were badly fit. One of the 2 galaxies with a bad fit (galaxy 66) is known to have strong emission lines.

Current evidence suggests that for about half of the DRG sample the IRAC MIR imaging improves the constraints on the ages and extinction levels. It is therefore imperative to see whether this has changed the median best-fit properties of the galaxies, most importantly estimates of their stellar masses.

4 Comparison to Lyman Break Galaxies

Our previous studies of the broadband SEDs (Franx et al. 2003; Förster Schreiber et al. 2004), and emission lines (van Dokkum et al. 2004) have indicated that at a given rest-frame optical luminosity DRGs are older, more obscured, and more massive than LBGs.

The results are still somewhat preliminary however, as even at optical wavelengths the effects of dust and on-going star formation introduce uncertainties. This may be particularly relevant when comparing the stellar masses of DRGs to those of galaxies with unobscured star formation, such as LBGs. Papovich, Dickinson, & Ferguson (2001) demonstrated that it is in principle possible to hide $5\times$ the stellar mass under the glare of active, unobscured star-formation, in a maximally old stellar population.

The IRAC mid-infrared imaging greatly reduces such uncertainties, as the rest-frame $K-$band light is a better indicator of the amount of stellar mass. In Figure 4 we present the result of stellar population modeling with the MIR fluxes included. The main result is that the median properties of both samples are unchanged with respect to earlier modeling based on the optical-NIR SEDs (Förster Schreiber et al. 2004). In particular, we find no evidence for an old and massive population

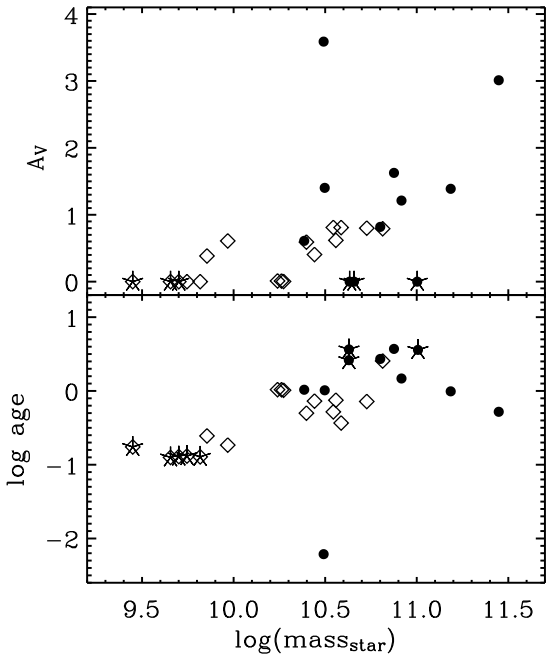

Figure 4 — (*a*) The best-fit values of Bruzual & Charlot (2003) model fits to the broadband SEDs of DRGs (*filled circles*) and Lyman Break Galaxies (*diamonds*). The stars indicate galaxies that are best-fit with an unreddened single age burst, others were fit with constant-star formation and Calzetti et al. (2000) dust. The DRGs are on average older, dustier, and more massive than LBGs.

hiding in the Lyman Break Galaxies. This is in line with IRAC studies of ground-based LBG samples Barmby et al. (2004), which are at somewhat higher redshift ($z \sim 3$) than our space-based U-dropout galaxies ($z \sim 2.5$).

Using the same fitting method as in §3 with the full complement of data, we find the following average values for the 11 out of 13 DRGs with good fits: a mean stellar mass of $< M_* >= 6.8 \times 10^{10} M_\odot$, a mean age of $< t >= 1.8$ Gyr, and a mean visual extinction $< A_V >= 1.7$, with a large dispersion in all properties. To calculate the averages, we used the biweight estimator (Beers et al. 1990), which is not sensitive to outliers. The 3 out 11 galaxies that were best-fit with an unreddened single age burst have an median age of $t = 2.6$ Gyr, the maximum age of the universe at $z \sim 2.5$. In contrast, the median properties of LBGs in the HDFS indicate that they are typically 4 times younger, are 4 times less obscured, and have stellar masses that are ~ 4 times lower.

Other systematic uncertainties on the mass estimates remain, such as the shape of the initial mass function (IMF) and the mass-limits, or the effects of metallicity, but these can only be resolved with direct spectroscopic measurements of kinematics and emission line ratios, which is difficult for DRGs (see e.g., van Dokkum et al. 2003, 2004).

Even without relying on SED modeling we can use simple arguments to estimate the relative contribution of two samples to the stellar mass density at $z \sim 3$. In the redshift range $2 < z < 3.5$ we find 11 DRGs and 32 LBGs to the same $K = 22.5$ magnitude limit. The DRGs emit about 60% of the light from LBGs at 8μ, but have redder $K - 8\mu$ colors, $< K - 8\mu >= 3.5$ versus $< K - 8\mu >= 3.1$ respectively. The $K - 8\mu$ color corresponds to the rest-frame $V - K$ at $z = 2.5$. Bruzual & Charlot (2003) stellar population models indicate that the redder rest-frame $V - K$ colors of the DRGs translate into mass-to-light ratios M/L_K that are twice as high. These values are fairly insensitive to the effects of dust or star formation history.

Hence at the high-mass end the $K-$selected DRGs contribute somewhat more to the $z \sim 3$ stellar mass density than LBGs, despite their lower number densities. This contribution increases if the sample were selected in the 8μm band, which is closer to a selection by stellar mass. To a limiting magnitude at 8μ of 18.2, we recover 10 out of 13 DRGs, and 12 out of 23 LBGs, boosting the contribution of DRGs. Hence, in a mass-selected sample at high mass the DRGs contribute $1.5 - 2\times$ more to the stellar-mass density than LBGs.

The major uncertainties in these estimates are the low number statistics and the variations in space density of the red galaxies due to large scale structure. We note that the HDFS is known to contain more objects with red observed $V - H$ colors compared the HDFN (Labbé et al. 2003), although our second deep field MS1054-03 shows a similar surface density in DRGs over a much larger $5' \times 5'$ field (see, e.g., van Dokkum et al. 2003; Förster Schreiber et al. 2004).

5 The rest-frame K-band mass-to-light ratio

We used the stellar population fits to the full SED to obtain estimates of the mass-to-light ratios M/L_K. We show the results in Figure 5. In (a) the M/L_K is plotted against extinction showing that the M/L_K ratios are not sensitive to the effects of dusts. This is expected as dust absorption at rest-frame K is small (cf., the visual mass-to-light ratios, Förster Schreiber et al. 2004).

A fundamental aspect of the rest-frame $K-$band mass-to-light ratios is that they increase with aging of the stellar population. At the same time, complex star formation histories, such as bursts, cause a spread in the luminosity weighted ages, and thus in the mass-to-light ratios. Interestingly, the DRGs have systematically higher M/L_K than LBGs, as expected from their redder rest-frame $V - K$ color,

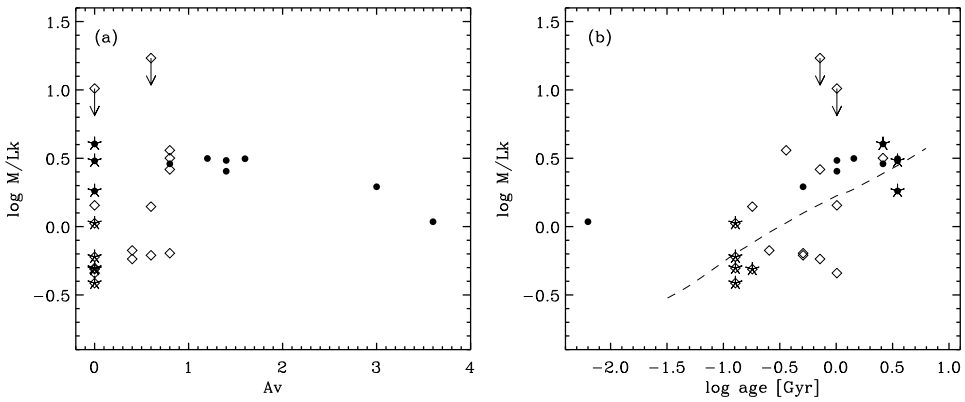

Figure 5 — The mass-to-light ratios in the rest-frame $K-$band (M/L_K) against visual extinction (a) and against age (b). The ages, masses and extinctions, were derived from Bruzual & Charlot (2003) stellar population fits to the SEDs. We used the best-fitting star formation history (single burst or constant) to calculate the masses. The galaxies best-fit with unreddened single age bursts are marked with a star. The DRGs (*filled circles*) and Lyman Break Galaxies (*diamonds*) are indicated. As expected, the observed $K-$band mass-to-light ratios are not sensitive to extinction effects, but do correlate with age. Overplotted in (b) is the evolution track of a Bruzual & Charlot (2003) stellar population model for constant star formation. Declining star formation histories follow similar tracks. The DRGs have higher mass-to-light ratios and less scatter in the mass-to-light ratios than LBGs.

but more importantly, their scatter is lower. Using the biweight estimator (Beers et al. 1990), we find $log(M/L_K) = 0.17(\pm 0.15)$ for the DRGs versus $log(M/L_K) = -0.22(\pm 0.45)$ for the LBGs.

This is consistent with a picture where DRGs are more evolved, and started forming at higher redshifts than LBGs. Similar conclusions were reached from emission line studies of a small sample of DRGs in the field of MS1054-03 (van Dokkum et al. 2004).

6 Discussion and Conclusions

The result presented here offer tantalizing insight into the nature of the red $J_s - K_s > 2.3$ galaxies at $z > 2$. The new mid-IR imaging with IRAC on Spitzer allowed us to uniquely identify "dead" galaxies with old passively evolving stellar populations. We placed better contraints on the relative role of age and dust for the star forming DRGs, breaking the degeneracy for some, and reducing it for many. Furthermore, we obtained more robust stellar masses, and mass-to-light ratios M/L_K in the rest-frame $K-$band.

We confirm the earlier studies of DRGs that found them to be more evolved, dustier, and more massive than LBGs (van Dokkum et al. 2004; Förster Schreiber et al. 2004). Specifically, we find average masses for the DRGs of $\sim 10^{11}$ $M_{\odot}$, and we excluded the existence of large amounts of stellar mass hiding under the glare of active star formation in our sample of $z \sim 2.5$ U-dropouts (cf., Papovich, Dickinson, & Ferguson 2001; see also Barmby et al. 2004). We find that for samples selected in the observed $K-$band, the DRGs contribute at least as much to the stellar mass density at $2 < z < 3.5$ as Lyman Break Galaxies. If the galaxies were selected in the rest-frame $K-$band, as a proxy for selection by stellar mass, then the DRGs would contribute $1.5 - 2\times$ more than LBGs.

The systematically higher mass-to-light ratios of the DRGs and the lower scatter have possibly far-reaching implications for scenarios of their formation and evolution. On the one hand, it strongly suggests that DRGs are more evolved and started forming at higher redshift than LBGs. They may have started out as Lyman Break Galaxies at $z \gtrsim 5$, and then endured a prolonged period of star formation which increased their stellar mass, metallicity, and dust content. This is consistent with studies of their emission line properties (van Dokkum et al. 2004). On the other hand, the comparable M/L_K, masses, and stellar ages of the dusty star forming and "dead" DRGs suggest that they are more closely related to each other than to LBGs.

While it can not be excluded that DRGs undergo a renewed "Lyman Break phase" after the addition of metal poor gas, the higher masses suggest that they are not simply LBGs seen along more obscured lines of sight.

How DRGs relate to lower-redshift galaxies is still unclear. Given their large masses, it is inevitable that they will evolve into massive galaxies locally, as they can never lose appreciable amounts of stellar mass. However, we cannot exclude that they evolve in complicated ways, and change their appearance after dramatic events, such as a gas-rich mergers (see e.g., Steinmetz & Navarro 2002).

Finally, we have found little evidence in the broadband SEDs of the DRGs that AGNs play a major role. We find a flux excess at 8 microns for 1 out of 13 galaxies, possibly related to an obscured AGN, although it did not affect the rest of the SED. For the LBG sample we find evidence for AGN activity at 8 microns in 3 out of 23 galaxies. The high fraction of AGNs found earlier by van Dokkum et al. (2004) in a spectroscopic sample of DRGss might have been a selection effect, or implies that most of the AGNs have low-luminosity. Chandra studies of DRGs in the FIRES MS1054-03 field suggest a luminous AGN fraction of 5% (Rubin et al. 2004), comparable to our findings. In this light it seems fair that we have interpreted our broadband SEDs exclusively in terms of stellar population properties.

The results presented here can clearly benefit from larger samples, and we are undertaking more studies at optical-to-MIR wavelengths towards this goal. In addition, large programs are underway to obtain follow-up optical and NIR

spectroscopy, which are constraining their space densities, and are providing independent ways to measure the extinctions, star formation rates, and masses. While such studies are now very hard, they will benefit tremendeously from the arrival of multi-object NIR spectrographs at $8 - 10$ meter class telescopes.

Acknowledgments

This research was supported by grants from the Netherlands Foundation for Research (NWO), the Leids Kerkhoven-Bosscha Fonds, the Lorentz Center, and the Smithsonian Institution. GR would like to acknowledge the support of the Deutsche Forschunggemeinschaft (DFG), SFB 375 (Astroteilchenphysik).

References

Barmby, P. et al. 2004, ApJS, in press
Beers, T. C., Flynn, K., & Gebhardt, K. 1990, AJ, 100, 32
Bolzonella, M., Miralles, J.-M., & Pelló, R. 2000, A&A, 363, 476
Bruzual, G. & Charlot, S. 2003, MNRAS, 344, 1000 (BC03)
Casertano, S. et al. AJ, 120, pp. 2747–2824, 2000
Calzetti, D., Armus, L., Bohlin, R. C., Kinney, A. L., Koornneef, J., & Storchi-Bergmann, T. 2000, ApJ, 533, 682
Efstathiou, G., et al. 2002, MNRAS, 330, L29
Fazio, G.G. et al., ApJS, in press
Franx, M. et al. *The Messenger* **99**, pp. 20–22, 2000
Franx, M. et al. 2003, ApJ, 587, L79
Förster Schreiber et al. 2004, ApJ, accepted (astro-ph/0408077)
Freedman, W. L., et al. 2001, ApJ, 553, 47
Giavalisco, M. & Dickinson, M. 2001, ApJ, 550, 177
Ivison, R. J., Smail, I., Barger, A. J., Kneib, J.-P., Blain, A. W., Owen, F. N., Kerr, T. H., & Cowie, L. L. 2000, MNRAS, 315, 209
Labbé, I., et al. 2003, AJ, 125, 1107
Labbé, I., et al, 2004, in preperation
Madau, P., Ferguson, H. C., Dickinson, M. E., Giavalisco, M., Steidel, C. C., & Fruchter, A. 1996, MNRAS, 283, 1388
Papovich, C., Dickinson, M., & Ferguson, H. C. 2001, ApJ, 559, 620
Rudnick, G. et al. 2001, AJ, 122, 2205
Rudnick, G., et al. 2003, ApJ, 599, 847
Rubin, K., et al. 2004, ApJ, accepted (astro-ph/0408101)
Smail, I., Ivison, R. J., Owen, F. N., Blain, A. W., & Kneib, J.-P. 2000, ApJ, 528, 612
Steidel, C. C., Giavalisco, M., Dickinson, M., & Adelberger, K. L. 1996, AJ, 112, 352
Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger
Steidel, C. C., Adelberger, K. L., Shapley, A. E., Pettini, M., Dickinson, M., & Giavalisco, M. 2003, ApJ, 592, 728
Steidel, C. C., Shapley, A. E., Pettini, M., Adelberger, K. L., Erb, D. K., Reddy, N. A., & Hunt, M. P. 2004, ApJ, 604, 534
Steinmetz, M. & Navarro, J. F. 2002, New Astronomy, 7, 155
Shapley, A. E., Steidel, C. C., Adelberger, K. L., Dickinson, M., Giavalisco, M., & Pettini,

M. 2001, ApJ, 562, 95
Shapley, A. E. et al. 2003, ApJ, accepted (astro-ph/0405187)
Spergel, D. N., et al. 2003, ApJS, 148, 175
van Dokkum, P. G. et al. 2003, ApJ, 587, L83
van Dokkum, P. G. et al. 2004, ApJ, accepted (astro-ph/0404471)